

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

lourinski, Dmitri (2008) A Dempster-Shafer theory inspired logic. PhD thesis, Middlesex University. [Thesis]

This version is available at: <https://eprints.mdx.ac.uk/6550/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

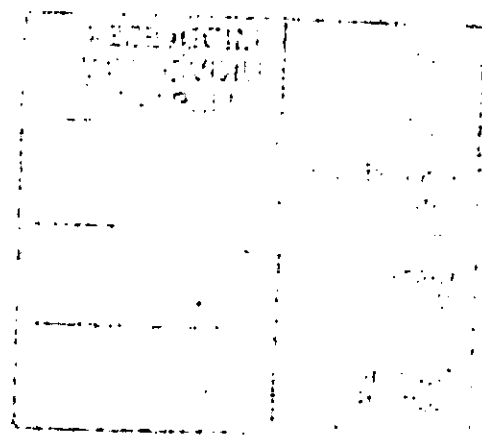
See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

A Dempster-Shafer theory inspired logic

Dmitri Iourinski

School of Engineering and Information Sciences
Middlesex University

A thesis submitted to Middlesex University in partial
fulfilment of the requirements for degree of Doctor of
Philosophy
November 2008



P1351954

Site HE	MIDDLESEX UNIVERSITY LIBRARY
Accession No.	0738002
Class No.	S11-3 100
Special Collection	

Dedicated to Melissa

Abstract

Issues of formalising and interpreting epistemic uncertainty have always played a prominent role in Artificial Intelligence. The Dempster-Shafer (DS) theory of partial beliefs is one of the most-well known formalisms to address the partial knowledge. Similarly to the DS theory, which is a generalisation of the classical probability theory, fuzzy logic provides an alternative reasoning apparatus as compared to Boolean logic.

Both theories are featured prominently within the Artificial Intelligence domain, but the unified framework accounting for all the aspects of imprecise knowledge is yet to be developed. Fuzzy logic apparatus is often used for reasoning based on vague information, and the beliefs are often processed with the aid of Boolean logic. The situation clearly calls for the development of a logic formalism targeted specifically for the needs of the theory of beliefs. Several frameworks exist based on interpreting epistemic uncertainty through an appropriately defined modal operator. There is an epistemic problem with this kind of frameworks: while addressing uncertain information, they also allow for non-constructive proofs, and in this sense the number of true statements within these frameworks is too large.

In this work, it is argued that an inferential apparatus for the theory of beliefs should follow premises of Brouwer's intuitionism. A logic refuting *tertium non datur* is constructed by defining a correspondence between the support functions representing beliefs in the DS theory and semantic models based on intuitionistic Kripke models with weighted nodes. Without additional constraints on the semantic models and without modal operators, the constructed logic is equivalent to the minimal intuitionistic logic. A number of possible constraints is considered resulting in additional axioms and making the proposed logic intermediate. Further analysis of the properties of the created framework shows that the approach preserves the Dempster-Shafer belief assignments and thus expresses modality through the belief assignments of the formulae within the developed logic.

Acknowledgements

I would like to thank my supervisors Dr. Roman Belavkin and Prof. Anthony White for their guidance throughout all stages of this research. I am also very grateful to my former supervisor Dr. Soodamani Ramalingam who introduced me to the Dempster-Shafer theory and without whom this work would not be possible.

I am indebted to my former senior colleague Prof. Vladik Kreinovich of the University of Texas at El Paso for introducing me to the field of artificial intelligence and to fuzzy mathematics in particular. The short, but fruitful collaboration with Dr. Francois Modave was another important milestone.

I am most grateful to my wife Melissa Bellovin without whose support and encouragement this work would never see the light of day. Her proofreading and editing help can not be measured or overestimated.

My parents, Ludmila and Vadim Jourinski were another invaluable source of support.

The other PhD students in the School were always good friends and gave me a very precious feeling of community that I will cherish for long time to come. Practically everybody contributed something positive, so I would not give a full list.

Contents

1	Introduction	1
1.1	The uncertainty representation	2
1.2	Fuzzy formalisms	3
1.2.1	Fuzzy sets	3
1.2.2	t -norms	5
1.3	The Dempster-Shafer theory interpretations	8
1.4	What follows	14
2	Selecting the framework	15
2.1	The theories in question	16
2.1.1	The Dempster-Shafer theory	16
2.1.2	Kripke models	18
2.2	The nature of mathematical objects	19
2.2.1	Historical Remarks	19
2.2.2	Phenomenological remarks	21
2.3	Implications on the formalism preference	23
2.3.1	Boolean logic	24
2.3.2	Modal logic	25
2.3.3	Intuitionistic logic	26
2.4	Modalities versus Beliefs	28
2.5	The formalism choice	29
3	Semantic models for the belief theory	31
3.1	Revisiting the theory of beliefs	31
3.1.1	The formalism	32
3.1.2	Calculating the beliefs	34
	The burglary of the sweetshop	36
	The cabbage seed	37

Alibi	39
3.2 Semantic models	41
3.2.1 Models and validity of formulae	41
3.2.2 Refuting <i>tertium non datur</i>	44
3.3 Translating frames of discernment to Kripke models	44
3.3.1 Constructing set W	45
3.3.2 A simplest frame	50
3.3.3 A slightly more complicated model	51
3.3.4 Updating the evidence	54
Homogeneous evidence combination	55
3.3.5 Building the models	56
Burglary of sweetshop	57
Cabbage seed	59
Alibi	61
3.4 Verification of the approach	62
3.5 Embedding issues	66
3.6 Incorporating the 'empty' evidence	68
4 Updating knowledge in the Dempster-Shafer theory	70
4.1 Refining	70
4.1.1 Refinement example	72
4.1.2 Refining versus evidence combination	74
4.1.3 Formalising model refinement	76
4.2 Properties of refinement	78
4.3 Coarsening	81
4.3.1 Coarsening and Kripke models	83
4.3.2 Coarsening example	85
4.3.3 Properties of coarsening Kripke models	86
4.3.4 Mass assignment and updating frames	87
4.4 Refining, coarsening and support functions	90
5 Discussion	95
5.1 The contribution	95
5.2 Comparison with earlier interpretations	97
5.3 Summary	99

List of publications

The results presented in this thesis have also appeared in the publications below.

1. D. Iourinski, "Phenomenology as criterion for formalism choice" in Rudolf Seising Ed. *Views on Fuzzy Sets and Systems from Different Perspective. Philosophy, Logic, Criticisms and Applications*, Springer, 2009
2. D. Iourinski, Dempster-Shafer theory inspired logic. In *Proceedings of 29th Linz seminar on fuzzy set theory*, Linz, Austria, 2008
3. D. Iourinski and R. Belavkin, Semantics of Dempster-Shafer inspired si-logic. In *Proceedings of 2nd IEEE-SOFA workshop on soft computing applications, Oradea-Romania, Gyula-Hungary*, 2007
4. D. Iourinski and J. X. For, Semantical Models for Evidential Frameworks. In *Proceedings of 2nd IEEE-SOFA workshop on soft computing applications, Oradea-Romania, Gyula-Hungary*, 2007
5. D. Iourinski, "A Dempster-Shafer theory inspired logic", in *Fuzzy sets and systems*, 2009 (Under revision)

Chapter 1

Introduction

This work presents an interpretation of the Dempster-Shafer theory [1] through a non-Boolean logic. From a pure mathematician's point of view it is an applied problem. From a computer scientist's viewpoint it is a theoretical work that draws new links between several related fields. In order to better appreciate the results achieved and the decisions taken along the way it is important to view this work in the wider context of the respective fields which inspired it. Even though the focus is on developing a logic formalism, the most important part of the work analyses the relationship and parallels between the proposed formalism and the evidential theory. In this sense the work is different from a dissertation in pure logic, such as [2] or [3]. The reasons for the choice of a logic are different from the ones of a pure mathematician. The author did not choose the logic that was the most interesting from the mathematical point of view. Instead, the logic that is the most meaningful from the Dempster-Shafer theory point of view was selected.

Semantic models or Kripke models are extensively used in this work for both representing and studying the evidential setups. From a computer scientist's point of view a Kripke model is first of all a directed graph. Within the domain, graphs are very attractive objects: they are easy to represent as a data structure, there are many efficient algorithms to manipulate them, the amount of formal knowledge about graphs is vast. Moreover, the earlier graphical representations of the Dempster-Shafer frames of discernment lacked generality and semantic analysis. Taking the above in consideration, makes providing a procedure that allows one to get both the graphical representation and the semantic interpretation of the Dempster-Shafer theory rather attractive.

The idea behind the whole undertaking is very simple. Boolean logic adequately represents the world described by classical probability theory, whereas the Dempster-

Shafer theory universe is different and merits its own inferential apparatus. There is no consensus about what this apparatus should be, but a large part of the proposed ones use modal logics (see the review below). Use of a modal operator for inference within a Dempster-Shafer universe contradicts the author's understanding of the nature of the beliefs, so a modality-free logic is proposed. Philosophical and mathematical factors influenced the decision. Both choices are justified in due course.

1.1 The uncertainty representation

The discussion about the nature of uncertain information is centred on two types of uncertainty: *aleatory* and *epistemic*. Aleatory uncertainty is the lack of knowledge about the system due to the system's random behaviour. Epistemic uncertainty stems from the lack of knowledge about the system that may well be operating according to some unknown non-random rules [4]. Aleatory uncertainty is traditionally addressed by probability theory, and there is little doubt about the theory's ability to do so [5]. The attention to epistemic uncertainty is more recent, and there is no consensus on the best approach. The review of the interpretations proposed so far is found in [4, 6, 7].

The Dempster-Shafer theory falls into a broader category of non-Bayesian statistics developed for solving different problems associated with the formalisation and interpretation of epistemic uncertainty. The decision making paradoxes are among the best known problems of this type [8]. The common feature of such approaches is their readiness to give up the additive probability measure on the universal set. What is used instead of the additive probability measure differs from one approach to another. The Dempster-Shafer theory abandons both the additivity of probability measure and the certainty of either event or its complement [9]. Schmeidler and later Wakker gave up the additivity and used fuzzy integrals for alternative ranking in decision under uncertainty [10, 11, 12].

Unlike many other theories, the Dempster-Shafer theory's beginnings are not difficult to trace. Dempster's article about upper and lower probabilities first appeared in 1967 [13], and the generalisation of the Bayesian rule in the new setup appeared in [14]. A few years later, Shafer developed Dempster's ideas into a full formalism in his doctoral dissertation [1]. It is interesting that the thesis presented a more general version of the theory than the monograph [9] that appeared later. The thesis addressed all possible setups, but the monograph limited its attention to the finite case only.

The Dempster-Shafer theory illustrates the modular nature of the formalisms for uncertain information. While the beliefs that replace the probabilities of the Bayesian setup are not necessarily additive, the sets which they represent are crisp. The Dempster-Shafer theory evidence combination rule relies heavily on Möbius transform [15], which is a combinatorial result only applicable to countable crisp sets. The original development used non-additive measures to describe crisp events. A fuzzy analogue of Möbius transform exists, but it was proposed later [16]. The evidence combination rule may be defined differently depending on the relationship among the sources of information [17].

1.2 Fuzzy formalisms

For the problem at hand it is convenient to think about fuzzy theory as a field that encompasses two distinct components. These fields are fuzzy measure theory and fuzzy logic. The two are clearly connected and one can, to a large extent, be developed from the other. In the setting of this work it is more convenient to look at them independently.

The author distinguishes between fuzzy logic in a narrow and a broad sense; the distinction is due to Hájek [18]. Fuzzy logic in a narrow sense is the logic that is constructed to formalise imprecise statements and thus facilitate formal reasoning. Fuzzy logic in a broad sense includes everything that deals with classes whose boundaries are not sharp. Building a non-Boolean inferential apparatus for the Dempster-Shafer theory is within the boundaries of the fuzzy logic in a narrow sense.

1.2.1 Fuzzy sets

The traces of questioning the bivalent logical statements are already present in Aristotle's works [19]. Computer scientists tend to relate the beginning of an alternative logic's development to the publication of Zadeh's article 'Fuzzy Sets' [20] in 1965. The article is a convenient reference point that served as a catalyst for the scientific community's interest in new ways to represent and utilise uncertainty. On the other hand, one of the most used fuzzy concepts, the Choquet integral, was introduced in the fifties without any reference to fuzzy sets. Choquet used the term 'capacities' instead, so the reference point is more or less arbitrary and reflects the birth of a term rather than of a theory.

The core idea of a continuous set membership function is quite intuitive and

fruitful: fuzzy set theory is a large field in itself that serves as a foundation for the development of new directions in the fields as divergent as set theory and topology [21, 22]. A good review of the fuzzy set theory and the corresponding issues can be found in [23]. Fuzzy sets as tools for epistemic uncertainty representation and interpretation are discussed in [24, 25]. Fuzzy sets are possibly the most popular fuzzy concept for applications. There are numerous applications in virtually any branch of computer science. A few works mentioned in the end of the paragraph can serve as a taster of how fuzzy sets are used in the field of pattern recognition: [26, 27, 28, 29, 30].

Given the new nature of the objects in question, one might expect that the classical measure theory would fail to address the needs of the fuzzy sets. However, the solution to the problem is more interesting than it looks at the first glance. A non-additive analogue of the Lebesgue's measure was first presented by Choquet in 'Theory of Capacities'[31]. Choquet also proposed a functional that could be used for calculating expected values of variables over sets characterised by *capacities* rather than *measures*. Zadeh later offered a more general version of Choquet's capacities which he called *fuzzy measures* [32], thus making Choquet's functional a fuzzy integral. Fuzzy measure is often interpreted as a generalisation of a familiar Lebesgue measure [33] and thus any kind of aggregation operator on it should reduce to the Lebesgue integral. This requirement is satisfied by Choquet's integral [34, 35] making it possible to develop a full fledged fuzzy integration theory. The fuzzy integration theory was developed by Sugeno [36], who also proposed his own aggregation min-max operator suitable for non-additive measures. This functional is called 'Sugeno integral', even though it is not an integral in the same sense as Choquet's integral: it does not reduce to the Lebesgue integral in the additive case.

The Choquet and Sugeno integrals are not the only possible aggregation functionals available for non-additive set functions. A number of alternatives were presented over the course of time, for example [37]. None of them, however, became as widespread as the two discussed above. The later advances in fuzzy integration theory include a unified framework called a *t*-conorm integral, which has both Choquet and Sugeno integrals as particular cases [38]. A structured review of the state of the art in the fuzzy integration theory in the mid-nineties is in [39].

1.2.2 t -norms

The inferential apparatus for dealing with imprecise information is based on two cornerstones — the t -norm and the modal operator. The idea of a multivalued logic is older than the idea of a fuzzy set. Multivalued logic and fuzzy logic are not synonymous. The agenda of the latter is wider: it is the theory that is meant to deal with all kinds of imprecise or vague information, while the multivalued logic deals with the partial truth of statements [40]. On the other hand, most of the theory available for multivalued logics is utilised to a different extent by the fuzzy community.

Some of the earliest formal developments in the field are due to Łukasiewicz, whose works first appeared in 1920-s [41, 42]. Łukasiewicz first introduced a three-valued logic that was later generalised to n -valued finite case. The logics with real-valued truth operators owe their existence to t -norms, a continuous or semi-continuous generalisation of the conjunction operator in classical logic. The definition of the t -norm is not very restrictive.

Definition 1.2.1 (t -norm) *A t -norm is a function $\star : [0, 1]^2 \rightarrow [0, 1]$ such that $\forall x, y, z \in [0, 1]$ it is*

1. *Commutative: $x \star y = y \star x$;*
2. *Associative: $(x \star y) \star z = x \star (y \star z)$;*
3. *Monotone: $x_1 \leq x_2$ implies $x_1 \star y \leq x_2 \star y$, and $y_1 \leq y_2$ implies $x \star y_1 \leq x \star y_2$;*
4. *Satisfies boundary conditions: $1 \star x = x$ and $0 \star x = 0$.*

The dual concept, a t -conorm is produced by calculating $1 - (1 - x) \star (1 - y)$.

The definition is not very restrictive and it makes many different t -norms possible which lead to various multivalued logics. The link between t -norms and multivalued logics is through a t -norm's residuum. It is known that the implication can be expressed as $\forall x, y \in [0, 1], x \rightarrow y = \max\{z \mid x \star z \leq y\}$ [18]. Research in t -norms is an active field, and many results, both theoretical and applied, are presented [43, 44, 45, 46]. Another important property of a t -norm is its algebraic connection: $[0, 1]$ equipped with a t -norm and its residuum forms a residuated lattice [47], thus allowing one to look for algebraic parallels when a logic is analysed. The algebraic parallels between logics and residuated lattices will be discussed in due course.

The second logic-building cornerstone is the modal operator. In some sense, introducing a new operator results in more dramatic changes to the system than a

different definition of the connectives. The effects of introducing the modal operator are discussed in more detail in Chapter 2. At the moment, it is enough to remember that the minimal effects of adding a modal operator to the Boolean propositional language include at least one new axiom (in addition to the axioms of Boolean logic) and a new inference rule [48]. This is not the case with a residuum of a t -norm: most multivalued logics use only modus ponens as the inference rule and operate within essentially the same propositional language as the Boolean logic.

The modal operator is not necessarily unique, as different modal logics may have different modal operators. The definition of modal operators and the set of axioms that they must satisfy depends on the area that a particular logic describes. There is no consensus among the researchers on what a modal operator is [49]. Among the well-known modal logics are the temporal logic [50]; the deontic logic that operates with concepts like obligation; the epistemic logic that formalises different types of knowledge with the modal operator 'is known that', and many others [51]. A global picture of the general modal logic and its various specific strands may be found in [48, 52, 53].

Even though introducing a modal operator is an effective logic building step that enables the construction of a bespoke inferential apparatus for different situations, the author argues against it. The details of the argument are presented in the next chapter. At the moment, it is enough to observe, that the modal operator in an inferential apparatus for the system of beliefs is 'excessive', because the modality is already expressed through the fuzzy measure of the subsets in the universal set. The purpose of the work is thus to construct an interpretation that allows a t -norm based logic construction, but not an introduction of new operators ('new' with respect to the propositional language of the Boolean logic). This approach can be called 'semantic-centred': instead of defining new operators or inferential rules, the author thinks of a logic as a set of true formulae and tries to understand which set gives an adequate description of the theory in question.

Another possible approach to creating the logics that are neither Boolean nor modal is through amending some already existing logic. The Boolean logic is the most established one, and its set of axioms is often used as yardstick when a new logic is analysed. Amending means either adding or removing some of the axioms from the set of the Boolean axioms. One of the earlier developments of this kind is due to Brouwer, whose constructive mathematics is based on rejecting the *tertium non datur* axiom: $p \vee (p \rightarrow \perp)$ [54, 55]. The resulting logic was called an *intuitionist logic*. Depending on the need, some extra axioms that are not *tertium non datur*

can be added, thus forming a superintuitionistic or, as the modern usage goes, an intermediate logic. The interest in these logics was first spurred by the attempts to create a new kind of mathematics, which does not allow a non-constructive proof and is generally stricter than the classical theory. Brouwer went further in his attempts to redevelop the mathematical notions. For example, he replaced a function with more general notions of a sequence and a fan [56, 57]. Not all of Brouwer's heritage is used in the present times. Moreover, the form in which Brouwer's ideas are known to the majority is due to his student Heyting, who published intuitionist works using the conventional notation [58]. Brouwer's intuitionism falls into a wider trend in mathematics called *constructive mathematics* [59]. The monograph [60] gives an account of Soviet constructive mathematics.

The new axioms can be presented based on two different premises. A choice of a t -norm puts certain restrictions on the set of true formulae. Several well-known logics such as Łukasiewics logic and product logic were created this way. Another approach is based on the requirement to meet certain semantic properties without any reference to a particular t -norm. Such an approach leads to the development of both two-valued and multivalued logics: the Medvedev logic of finite problems and many others were created this way. The intermediate logics are not necessarily finitely axiomatizable. Moreover, the logic proposed in this work is not finitely axiomatizable, but has a simple definition based on a first-order condition on the underlying semantical structures. The field is just too large to give a detailed overview. A good overview of important intermediate logics is in [48, 61, 62], the rest of references is given as the need arises.

This work explores the properties of the intermediate logics using the well-known correspondence between logics and lattices. The primary tool for semantics exploration is a Kripke model. Viewing logic as algebra is not unique for modal or multivalued logic. Many completeness results within Boolean logic owe their existence to the algebraic representation and parallels. An 'algebraic' view of the familiar Boolean logic can be found in [63]. A good summary of the relevant results is in already cited [48, 61, 64] as well as [65, 66]. The algebraic view of logic falls within *metamathematical* view of the subject. The classical exposition to metamathematics is in [67, 68].

While the algebraic approach is applicable to a very wide spectrum of logic problems, the semantic models proposed by Kripke are not equally universal. Although the classical Boolean logic may be represented using a Kripke model, the model is degenerate and thus not interesting. The situation changes dramatically when in-

intermediate or modal logics are involved. Using Kripke models allows for an easier translation to the language of lattices, and it helps to use combinatorial results for the semantic analysis among other uses. The construction presented in this work also relies on Kripke models, so the discussion of the particular merits of the approach is postponed until the relevant notation is introduced and the results are proved. It is also worth mentioning that originally the Kripke models were introduced as a tool for the philosophers. A detailed analysis of Kripke's philosophy can be found either in his own works, [69] or in monograph [70].

1.3 The Dempster-Shafer theory interpretations

This work is synthetic in its nature: it contributes to the decision making theory by constructing a procedure that maps the Dempster-Shafer belief theory to the semantic models and then uses algebraic methods to analyse the semantics of the result. The idea is quite natural: while the Boolean logic provides a perfectly adequate inferential apparatus for the Bayesian statistics, it is not the case with the belief theory. There are several challenges to be addressed. A 'useful' interpretation should be general enough to translate any possible Dempster-Shafer universe to a semantic model, yet flexible enough to allow for later evidence updates through both the Dempster-Shafer evidence combination rule and frame transformations. Not all multivalued logics can do both things. For example, the logic in [71] cannot represent the Dempster-Shafer evidence combination rule. Aside from the formal considerations, different interpretations of the meaning of beliefs are possible. The discussion on the topic is quite fruitful, as in [40, 72], which offer a more general discussion on the requirements to such interpretation. The discussion about different interpretations of the Dempster-Shafer theory inevitably addresses the question of processing uncertainty and the place of logics in it.

According to Klir [7] the uncertainty is processed on three levels: Formalisation, Measurement and Utilisation. The Dempster-Shafer theory is a formalisation technique, logic is used for the utilisation of uncertain information. Logic interpretations of the Dempster-Shafer theory always make decisions about the relationship between different levels of processing uncertainty. Often these decision are based on understanding uncertainty as modality.

Even though the rigorous discussion about choosing the appropriate set of tools is postponed until the next chapter, the overview of the existing interpretations is not quite possible without reference to a few technical concepts. The modal logic is

commonly defined as an extension of the Boolean logic, which among other things includes *modal connectives*: \Box (necessity) and \Diamond (possibility). The formal properties of these operators may differ depending on the logic and will be introduced as the need arises. Another concept used by most non-classical logics, is a *possible world*. While the Boolean logic operates with a static world, non-standard logics often allow for different states of affairs (worlds) in which different variables are instantiated. The collection of possible worlds, along with accessibility relation and function that links the worlds and formulae are referred to as *semantic or Kripke models*.

One of the earlier developments in the field was the possibilistic logic proposed by Dubois and Prade [73]. It is a logic of weighted formulae that is in agreement with Zadeh's understanding of necessity and possibility measures. The possibilistic logic was not developed specifically for the needs of the Dempster-Shafer theory. Dubois et al. presented a framework that departs from the truth functionality of the fuzzy logic thus providing a general approach for the researchers who do not view uncertain propositions as truth-functional. Following the same philosophy Farinas and Herzig [74] proposed a qualitative possibility logic, in which they axiomatize the notion of qualitative possibility based on ordering possibilities and necessities of propositions.

Another logic in this family is due to Boutilier [75] who developed two possibilistic logics for reasoning under uncertainty. Boutilier's results include two logics CO and CO*. Both constructions are extensions of Boolean logic. Even though the resulting constructions are modal logics, the approach is different from the majority of modal logics that appear in the literature. There are two modal operators. The first modal operator is familiar \Box (modal necessity) that stands for truth in accessible worlds. The other operator is unusual \Box^\perp describing the truth in inaccessible worlds. The semantics is based on Kripke models for two valued logics. The uncertainty of propositions is described through the possibility and necessity measures. The meaning of a possibility measure is interpreted in terms of the amount of surprise associated with a statement. The higher the possibility measure of a statement the lower is the observer's surprise when the statement is true.

Ultimately, Boutilier aims to develop a qualitative reasoning apparatus and a large part of work is devoted to going from quantitative notions to the qualitative ones. Qualitative ordering is possible and given. There is a distinction between epistemic possibility and physically or logically possible worlds that an agent can possibly consider adopting.

The possible world semantics is used, the accessibility relation R is understood as ranking according to the degree of possibility, R is reflexive and transitive preorder.

R is a total preorder: two states of affairs must be comparable according to their degree of possibility. wRv means that v is at least as possible as w . In other words, the accessibility ranking gives ordering from least possible to the most possible world. Language is a countable set P of propositional variables along with connectives $\neg, \supset, \Box, \Box$. Semantic CO models in this case are $M = \langle W, R, V \rangle$, where R is transitive, connected binary relation on set of possible worlds W and V is a multivalued map between P and W . The semantic models also contain clusters of worlds ($vRw \wedge wRv$) that are always present. The clusters can be thought of as the sets of worlds that are equally possible. The truth of formulae is defined through the modelling relation

1. $M \models_w \Box \alpha$ iff for each v such that wRv , $M \models_v \alpha$.
2. $M \models_w \Box \alpha$ iff for each v such that vRw , $M \models_v \alpha$.

The connective \Box is then the truth in the worlds that are at least as probable as w , and connective \Box is the truth in all inaccessible worlds, i.e. the ones that are less possible than w .

The logic CO* is the smallest extension that assigns positive degrees of possibility to every logically possible world. CO* is closed under rules of CO and has an extra axiom, a model in this logic is then any CO-model such that $\{f : f \text{ maps } P \text{ into } \{0, 1\}\} \subseteq \{w* : w \in W\}$

In order to have qualitative reasoning but not to lose in expressive power, Boutilier introduces quite a few non-standard logical connectives and ultimately gives a set of axioms. The final product is a logic that includes Boolean logic, several additional axioms and is closed under necessitation and modus ponens. Overall, the approach produces an ordering of the formulae but no truth values. The main idea is similar to the approach taken in this work except that there is a modality and things are not quantified, but instead ordered using relation R . The drawback of the approach is that it does not allow for an analogue of the Dempster-Shafer evidence combination rule, as there are no numerical possibilities.

The logics reviewed above used two valued connectives. Using multivalued connectives for the Dempster-Shafer theory interpretation was also done. One of the better known attempts is due to Hájek et al. [76] who proposed an extension of the Boolean logic by introducing the beliefs on Boolean formulae. The belief in this interpretation is understood as a truth degree of a fuzzy proposition $B\varphi$ which stands for ' φ is believed'. The belief is defined as probability of modal necessity. A set of new axioms is introduced. The result is a combination of logic S5 with fuzzy approach named $L\Pi_2^1$ — a combination of product and Lukasiewicz logics.

The notion of probability is generalised by taking into consideration expressions like

“belief degree of φ ” = “truth degree of $B\varphi$ ”,
 where $B\varphi$ is a fuzzy proposition “ B is believed”.

The semantic of such logic is studied with aid of \square -probabilistic Kripke models: $\mathcal{K} = \langle M, \mu \rangle$, where M is a \square -Kripke model (modal Kripke model). A *modal Kripke model* is defined as $\mathcal{M} = \langle W, R, V \rangle$ where R is an equivalence relation, and valuation function V is extended to modalities. μ in a \square -probabilistic Kripke model is a finitely additive probability measure over the algebra of subsets of W . Beliefs in this setting become the probabilities of necessities. It is shown that beliefs can be represented by some Kripke model.

Logic $\text{L}\Pi_{\frac{1}{2}}$ is a logic built with constants $\bar{0}, \frac{1}{2}$ and $\bar{1}$ using connectives $\rightarrow_L, \rightarrow_{\Pi}$ and \odot (Łukasiewicz and Goguen implications and product conjunction). There are many other connectives defined through the three already mentioned operations. The evaluations are given through values of function e : $\text{For}(\text{L}\Pi_{\frac{1}{2}}) \mapsto \{0, \frac{1}{2}, 1\}$. Finally, $\text{L}\Pi_{\frac{1}{2}}$ is used to define a logic that includes both Boolean and multivalued formulae and has a probability operator. As a result, the notions of tautology and provability are redefined. The final axiom set for the logic includes modal logic S5, provable formulae in S5, axioms for fuzzy notion of being probable and $\text{L}\Pi_{\frac{1}{2}}$ axioms.

Overall the construction is powerful and displays good completeness properties. Unlike the previous construction, there is some analogue of Dempster-Shafer evidence combination rule. Given the richness of the construction, combining different models in it becomes quite complicated. Obtaining new probability assignment now involves combining three pairs of modalities.

The two logics above were built for the case $\text{bel}(\emptyset) = 0$. Replacing S5 with a weaker KB4 logic makes it possible to repeat similar construction for belief functions allowing $\text{bel}(\emptyset) > 0$.

Boeva, Tsiropkova and De Baets [77, 78] take the approach that is the most close to the approach taken in this work. The authors attempt to build the minimal modal logic for the needs of the Dempster-Shafer theory. Boeva et al. depart from the model in which value assignment function requires exactly one proposition to be true at each world. Instead, they allow for an arbitrary number of propositions to be true in each possible world. The plausibility and belief measures are induced by the accessibility relation which they view as a multivalued mapping from set of formulae to sets of worlds.

The multivalued mapping is a map $F : X \rightarrow Y$ that assigns a subset $F(x) \subseteq Y$

to every $x \in X$. The domain of a multivalued mapping is given by $\text{dom}(F) = \{x \in X : F(x) \neq \emptyset\}$. Subsets of Y then may have differently defined *inverse images*, among which the inverse, the superinverse and the pure inverse are distinguished. The multivalued mappings and probability measures can be related. Consider a probability measure P on $\mathcal{P}(X)$. It is then possible to find a multivalued mapping Γ , such that $P(\text{dom}(\Gamma)) > 0$. Dempster's upper and lower probabilities are then interpreted as probability measures of inverse and superinverse images of different subsets within domain Γ . The basic probability assignments can also be expressed through the inverse images of the sets within domain Γ .

The definition of the modal logic used by the authors is standard: a set of atomic propositions, logical connectives $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$, and modal operators of possibility \Diamond and necessity \Box . The modal logic is analyzed through the semantic models $\mathfrak{M} = \langle W, R, V \rangle$. The modality is understood through the accessibility relation: \Diamond is equivalent to being in the truth set of a proposition, and \Box is equivalent to being able to see elements from a truth set.

Viewing the accessibility relation R as a multivalued mapping allows one to find the plausibility and belief measures of propositions. Such approach induces plausibility and belief measure and basic probability assignments on any model using a probability measure on set $\mathcal{P}(W)$ as a starting point. The resulting models are reasonably well-behaved and satisfy Weak Singleton Valuation Assumption requiring that at least one proposition is true in each world, thus giving some protection against vacuous reasoning.

The weak singleton valuation assumption is important in the context of the current work. The models that are built upon the same premise are used to generate the intermediate logic.

In order to be a full-fledged representation of the Dempster-Shafer theory, the models should also be able to represent the Dempster-Shafer evidence combination rule. The proposed procedure is quite simple: two models $\langle W_1, R_1, V_1, P_1 \rangle$ and $\langle W_2, R_2, V_2, P_2 \rangle$ are combined by taking Cartesian products of corresponding sets and the new probability measure is the product of corresponding measures. Regardless of the simplicity of the procedure, it results in a new model which is equivalent to the orthogonal sum of its components.

The authors of the approach did not analyze the resulting class of semantic models and did not analyze the corresponding modal logic as a set of true formulae. The proposed approach is different from most other modal logic interpretations of the Dempster-Shafer theory in a sense that it does not introduce any requirements on

the modality, but instead uses the most general definition of it and then builds a logic using only the most fundamental premises. In the next chapter this argument is taken even further. By arguing in favor of developing an interpretation that bypasses the modality altogether and attempts to express the possibilistic reasoning through the means of an intermediate logic.

The logics overview of which was just given share one fundamental thing in common: they start with some understanding of uncertainty as modality and build their models from there. For the sake of brevity, this approach is called *operator-centred*. The starting point is a specific definition of the modal operator. This definition is based on an author's intuition about the nature of the beliefs. The resulting logic is then to a certain degree determined by the choice of the operator in the propositional language. The approach in this work is different, as it attempts to translate the Dempster-Shafer theory beliefs to the Kripke models and then to analyse the semantics of the resulting models. The operators are chosen from those that can express the created semantics. Using a modal operator was ruled out from the very beginning. The detailed justification of this choice and the discussion about its consequences are in Chapter 2. However, even if the modality were introduced, the choice of possible logics would still be different from the ones discussed above.

The approach taken in this work is similar in spirit to one of the earliest graphical interpretations of the Dempster-Shafer theory due to Barnett [79] who first proposed a linear time algorithm for calculating the beliefs. The proposed algorithm worked with only one particular type of belief functions producing binary trees. Later improvements on the technique include [80] and a work by Shafer and Logan [81]. Guan and Bell took the method to its logical conclusion giving both the formal description and the corresponding algorithms [82]. The works mentioned above do not give a semantic representation of the theory; their purpose is quite different. All of the works represent the Dempster-Shafer universe with the aid of trees, using them to calculate the beliefs without venturing into semantics.

To this point, a very brief overview of the main relevant theories and approaches was given. The main coalescence points that, to a certain extent, served as the guiding markers in the development of the work's approach have also been presented. Further discussion on the background and available results is impossible without introducing rigorous definitions and quoting actual results. Any further references to published results will be given as need arises and only the notions needed for understanding new results will be given. The next chapter addresses the question of a formalism choice for the interpretation.

1.4 What follows

The next chapter is devoted to the discussion about the nature of mathematical objects. The purpose of the chapter is to demonstrate how the philosophical considerations may affect the formal choice when an interpretation is constructed. After reviewing the relevant theories, the argument in favour of following the intuitionist view of the world is given and the decision to create an intermediate logic is made. The material of this chapter is due to appear in [83].

Chapter 3 gives the necessary background about the Dempster-Shafer theory and semantic models and presents the procedure that links them. After analysing several examples, it is then proved that the procedure preserves the beliefs of the original frame of discernment. Similarly it is shown that the evidence combination rule is translated to the language of the semantic models [84]. Once the approach is verified, some attention is paid to the resulting superintuitionistic logic which is shown to be sound and complete. The completeness results are produced by analysing semantically equivalent algebraic duals of Kripke frames generated by application of the procedure proposed earlier in the chapter.

Chapter 4 explores the parallels between frame updates and semantic models. While studying the effects of frame coarsening and refinement on a semantic model, it demonstrates the representational limitations of the approach. The last chapter gives a brief overview of open questions and possible directions of research.

Chapter 2

Selecting the framework

In the previous chapter several different interpretations of the Dempster-Shafer theory based on modal logics were mentioned. The major problem with these formalisms is, in the author's view, the possibility of a non-constructive proof. The current work does not present yet another logic that works relatively well, but attempts to understand how to construct a family of logics that do not allow non-constructive proofs and can be used for inference within the realm of the Dempster-Shafer theory.

The author takes advantage of the 'modular design' of building a logic in a propositional language. As long as there is *some* definition of a propositional language, one can explore the semantics of logics built upon it. The logical connectives can be defined later. For the problem at hand the choice of suitable logical connectives is limited by the known semantic limitations. Most other publications on the matter approach the problem from the other end — either the connectives are chosen first, or the axioms are stated explicitly before the logic is being built.

While having many merits of its own, defining a logic explicitly through a set of axioms rules out all of the logics that do not admit finite axiomatisations. On the other hand, it is known that many otherwise well-behaved logics do not admit finite axiomatisation. Specifying a Hilbert-style calculus is not the only way to represent a logic, especially within the computing domain. The discussion below is not very technical, as the formal considerations are postponed until the next chapter. It is, however, impossible to choose a formalism without resorting to at least some formal concepts, which are given below.

2.1 The theories in question

The main topic of this work is the justification of the formalism choice for interpreting the Dempster-Shafer theory. The approach taken by the author is focused on understanding the semantic of the constructed formalism. The semantic tool of choice is Kripke models. There is a vast amount of literature about both Kripke semantics and the Dempster-Shafer theory. The Dempster-Shafer theory overview below is based on Shafer's original essay [1]. The semantic models are defined according to [48]. The definition of a Kripke model used in this work is slightly different from the one used in important works like [18], but given the popularity of the concept it is quite difficult to decide on a particular notational convention. An interesting review of different definitions of Kripke model is given in [85]. A more detailed introduction to both theories is given when the formal aspects of representation are discussed. The definitions below are solely meant to make the epistemological discussion consistent.

2.1.1 The Dempster-Shafer theory

There is no complete specification of the propositional language used in the Dempster-Shafer theory. Instead, there are several conditions that the statements of this language should meet. The propositions are related to subsets of a given set. Let θ be a quantity of interest and Θ be the range of its values. The possible propositions are of the form

$$\text{"The true value of } \theta \text{ is in } T\text{"}$$

where $T \subseteq \Theta$. The universe formed this way is not as restrictive as it may seem, and all the possible statements are in one-to-one correspondence with subsets of Θ . In the Dempster-Shafer theory range of values Θ and its known subsets are called a *frame of discernment*.

A frame of discernment does not include actual propositions. It describes the domain of the values that quantities in the propositions can assume. Therefore the propositions considered in the Dempster-Shafer theory form a propositional language \mathcal{L} , which is not yet defined. \mathcal{L} can be defined when the possible candidate logics are considered. A good discussion about how to define a propositional language is found in [63, 86].

That said, intuitive guesses can be made about the nature of the relation between

\mathcal{L} and Θ . Indeed, assume that $p, q \in \mathcal{L}$; $A, B \subseteq \Theta$ and that p and q stand for:

$p = \text{"the true value of } \theta \text{ is in } A\text{"}$

$q = \text{"the true value of } \theta \text{ is in } B\text{"}$

It is not illogical then to assume that

$p \wedge q = \text{"the true value of } \theta \text{ is in } A \cap B\text{"}$

The correspondence between \wedge in \mathcal{L} and \cap in Θ is not necessarily true for all the frames of discernment and all the languages, but it is very intuitive. The correspondence is true when \mathcal{L} is the language of Boolean propositional calculus — the situation considered in Shafer's monograph.

A *belief function* is defined on the frame of discernment through the set of requirements.

Definition 2.1.1 (Belief function) Let Θ be a frame of discernment and $Bel : 2^\Theta \rightarrow [0, 1]$ be a set function such that

1. $Bel(\emptyset) = 0$;
2. $Bel(\Theta) = 1$;
3. For every positive integer n and every collection A_1, \dots, A_n of subsets of Θ ,

$$Bel(A_1 \cup \dots \cup A_n) \geq \sum_{I \subseteq \{1, \dots, n\}; I \neq \emptyset} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right).$$

Bel is then a belief function on Θ .

The Dempster-Shafer theory also uses the concept of a *basic probability assignment*.

Definition 2.1.2 (Basic probability assignment) Quantity $m(A)$ is called a *basic probability number (assignment in newer works)* if it obeys the following restrictions:

1. $m(\emptyset) = 0$;
2. $\sum_{A \subseteq \Theta} m(A) = 1$.

$m(A)$ measures the belief that is committed exactly to A . A belief function and a basic probability assignment are related through:

$$Bel(A) = \sum_{B \subset A} m(B). \quad (2.1)$$

The equation above serves as an alternative definition of a belief function.

Naturally, a logical interpretation of Dempster-Shafer theory must preserve the properties above. The definitions demonstrate the dual nature of the basic probability assignments and the belief functions. The mathematical apparatus dealing with this duality allows conversion in either direction.

Evidence is updated by introducing new propositions with their own mass assignments. The new pieces of evidence transform already established beliefs. Two different frames of discernment within the same universe may be combined too. Combining new and existing information is done by taking the orthogonal sum of the respective mass assignments. Formal definition of the orthogonal sum is not yet needed and at the moment it is enough to know that there is a meaningful interpretation of it in the proposed formalism. The goal at the moment is to find a logic that can adequately represent frames of discernment.

2.1.2 Kripke models

Kripke models are a famous tool for exploring different modal logics. It is important to remember that Kripke models are not exclusively applicable to the modal logics. They provide a formalism for addressing a wider range of objects.

Definition 2.1.3 (Kripke Model) *Given propositional language \mathcal{L} , intuitionistic Kripke model is a triple $\mathfrak{M} = \langle W, R, V \rangle$, where W is a set, R is a partial order on W , and $V : \text{Var}\mathcal{L} \rightarrow \text{Up}W \subseteq 2^W$ is a valuation map, where $\text{Up}W$ is the collection of all upward closed subsets of poset (W, R) : $x \in M \subseteq W$ and xRy imply $y \in M$ for all $M \in \text{Up}W$. Together R and V satisfy the property of persistence of propositional variables: $\forall v \forall w \forall p (vRw \rightarrow (v \in V(p) \rightarrow w \in V(p)))$.*

The elements of W are sometimes called *possible worlds* or, less dramatically, *points*. xRy , $x, y \in W$ is read either ' x sees y ' or ' y is reachable from x '. The definition is not very restrictive and leaves a lot of space for the manoeuvre. A usable interpretation of the Dempster-Shafer theory may be developed by an intuitive understanding of the universe described through Kripke models. The elements of W can be thought of as different states of information or knowledge. The valuation V provides the link between the actual knowledge (the propositions of \mathcal{L}) and the states

of knowledge (points of W). Different statements are true in different states. The relation R shows what could be inferred from different states of knowledge. If point x sees point y , it means that the information available at y may be inferred from information available at x . Point x occurs *earlier* than point y . If a proposition is true at point x , it cannot become false at later points reachable from x . Conversely, a proposition false at some point can become true at a later point reflecting the ability to discover new facts. Graphically Kripke models are represented as directed graphs with vertex set W and adjacency matrix given by relation R (see [87] for a definition of the adjacency matrix).

2.2 The nature of mathematical objects

The Kripke models can be used to analyse the semantics of any logic and they provide alternative definitions of logics in certain cases. The idea of the proposed approach is to develop a procedure that allows an adequate representation of the frames of discernment by Kripke models using *some* propositional language and thus to induce a family of logics whose formulae are valid in the corresponding Kripke frames. The strategy does not necessarily lead to a unique solution: depending on the choice of the propositional language the same models can correspond to different logics. The search domain can be narrowed by looking at the philosophical and consequent semantic distinctions among the three main strands of logics: Boolean logic, modal logic and intuitionistic logic. The exposition starts with some phenomenological remarks that are later applied to the ‘candidate logics’. Even though there is no ‘single best’ logic for inference within the Dempster-Shafer theory, the discussion outlines the main arguments in favour of using intuitionistic logic or its non-classical extensions, as well as author’s motivation for not introducing modal operators.

2.2.1 Historical Remarks

This work is devoted to the development of a mathematical formalism suitable for reasoning within the Dempster-Shafer theory. Thus, the subject of enquiry is a collection of mathematical objects. In the belief theory the ability to update the knowledge, possibly as a result of the interaction with the outside world, is important. The constructed collection of the mathematical objects should not be independent from the notion of time. Discussing the temporality of mathematical objects cannot be done without a short foray into the philosophy of mathematics. The definitions of a

mathematical object and its basic properties are important for the decision made in this chapter, so the exposition starts with quoting the relevant views. The outline of the main philosophical issues is by no means comprehensive: only the issues that the author considers relevant for the problem at hand are mentioned.

While the discussion of the nature of mathematical objects is ages old, this work confines itself to the formalisms that date back to the XIX and XX centuries. Boole was the first person to introduce an example of a non-numerical algebra and the first example of a symbolic logic [88] thus paving the way for symbolic mathematics as we know it. The familiar Boolean logic in the modern notation is, however, due to Frege [89].

The Russell-Whitehead 'Principia Mathematica' [90] was published in 1910-3. It famously tried to develop all mathematical truths from a well-defined set of axioms and inference rules in symbolic (Boolean) logic. As the authors of [91] put it, Boole wanted to study the mathematics of logic whereas Russell and Frege wanted to study the logic of mathematics.

The Lindenbaum-Tarski approach provides a way to construct an algebra out of a classical propositional calculus. The constructed algebra is a distributive lattice. Moreover, the method works with both the classical propositional calculus and with any algebra formed by a closed (according to *some* definition) set of formulae in any propositional language [68]. As it will become clear later the latter fact is very important for the procedure being developed.

Wittgenstein published his 'Tractatus Logico-Philosophicus' in 1921 [92]. While agreeing with Russell at certain points Wittgenstein introduces a different understanding of what a mathematical object is and what is the purpose of mathematics and philosophy. In 1918 Brouwer begins the systematic intuitionistic reconstruction of mathematics with the paper '*Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil, Allgemeine Mengenlehre.*' ('Founding Set Theory Independently of the Principle of the Excluded Middle. Part One, General Set Theory') [93].

Around the same time, Husserl developed his phenomenological approach to mathematics [94]. A detailed account of the similarities and differences of the philosophical approaches can be found in [95]. This work only looks at a few basic distinctions between the approaches that are relevant to the development of the proposed formalism.

2.2.2 Phenomenological remarks

In addressing the stated goal of the work, the author follows several basic definitions. First of all, Wittgenstein's definition of the world captures the basic idea of artificial intelligence (quoted from [92]).

1.13 The facts in logical space are the world.

2.034 The structure of a fact consists of the structures of states of affairs.

2.04 The totality of existing states of affairs is the world.

The next important distinction is about the nature of the states of affairs that in our case are represented by the mathematical objects. One of the main distinctions is in the relation between the mathematical objects and time. Several dichotomies are possible [95]:

- *Static/dynamic*: an object is static exactly if at no moment are parts added to it, or removed from it. It is dynamic if at some moment there are parts added to it, or removed from it.
- *Temporal/atemporal*: an object is temporal exactly if it exists in time, and atemporal exactly if it does not exist in time.
- *Intratemporal/omnitemporal*: a temporal object is omnitemporal exactly if it is static and exists at every moment. A temporal object is intratemporal exactly if it is not omnitemporal.

A decision making formalism representing some model of the real world is not static, and thus the distinction between omnitemporal and intratemporal becomes important. Van Atten presents three logical possibilities [95]:

1. All mathematical objects are omnitemporal. (Husserl)
2. No mathematical objects are omnitemporal. (Brouwer)
3. Some mathematical objects are omnitemporal, some are not.

One of the fundamental premises of the theory of beliefs is the possibility to learn and incorporate new knowledge into the already known. The author also wants some uniformity of the objects, so the first and the third views are ruled out. Brouwer's view is the most attractive for the stated purpose.

Brouwer's philosophical views led him to the development of his own system of mathematical foundations that he called intuitionist mathematics. Brouwer further elaborates on his view of mathematical objects:

In intuitionist mathematics a mathematical entity is not necessarily pre-determinate, and may, in its state of free growth, at some time acquire a property it did not possess before [95].

Intuitionism is often viewed within a broader constructivist approach to mathematics. Constructivists, however, need not accept the idea of dynamic objects. The objects of interest are dynamic, hence the outlook may be restricted to intuitionist mathematics.

The concern with the notion of time is not unique for the intuitionism. Temporal logic is one of the most well-known and developed examples of the approaches that explicitly incorporate the notion of time into mathematical objects. On a more general level, temporal logics are a class of modal logics in which the modality expresses temporal relations. These modal logics are the obvious candidates for the reasoning apparatus for the Dempster-Shafer theory and the selection process starts with looking at their behaviour.

Even though it is already established that the objects of interest are dynamic and intratemporal there is more than one choice to be made. In a nutshell, one has to decide what is primary: the concept of the flow of time or the concept of change [50].

The author believes that the nature of the objects described by the Dempster-Shafer theory is better described through the approach when the notion of change is accepted as primary. Accepting primacy of the flow of time is the stand taken by the temporal logic. If the flow of time is primary, then the propositions hold truth values for some time and may change them as time passes. In intuitionistic Kripke semantic models such a situation is not quite possible: a variable can be instantiated at some moment of time, but it cannot change its value at a later time. Instead, accepting the primacy of the change transforms the moments of time into the equivalence classes, the situation rendered through the concept of the state of the world.

Temporal logic is then not a suitable candidate under the given premises. It does not mean that temporal logic cannot provide reasoning tools for the Dempster-Shafer theory: the successful applications of a temporal logic were discussed in Section 1.3.

Brouwer separates mathematics into old formalism; pre-intuitionism of Borel, Lebesgue and Poincaré; and new formalism. The intuitionism has two acts.

First act of intuitionism completely separates mathematics from the phenomena of language described by theoretical logic, recognising that intuitionist mathematics is an essentially language less activity of the mind having its origin in the perception of a move of time.

Second act of intuitionism admits two ways of creating new mathematical entities: firstly in the shape of more or less freely proceeding infinite sequences of mathematical entities previously acquired (e.g. infinite decimal fractions); secondly in the shape of mathematical species, i.e. properties supposable for the mathematical entities previously acquired, satisfying the condition that if they hold for a certain mathematical entity, they also hold for all mathematical entities which have been defined to be 'equal' to it, definitions of equality having to satisfy the conditions of symmetry, reflexivity and transitivity [56].

While commenting on the first act of intuitionism Brouwer introduces the notion of a *fleeing property* f :

- (i) for each natural number n it can be decided whether or not n possesses the property f ;
- (ii) no way of calculating a natural number n possessing f is known;
- (iii) the assumption that at least one natural number possesses f is not known to be an absurdity.

The acceptance of the fleeing property leads to the rejection of the *tertium non datur* principle and raises the problem of interpreting and intuiting the continuum, solved by the second act. At the same time, the second act weakens the restrictions of the first act: while for any proposition p it is known that $p \vee (p \rightarrow \perp)$ is true only if p is decidable, it follows from the second act that $\neg p \vee \neg\neg p$ is provable (absurdity or absurdity of absurdity in Brouwer's words).

In phenomenological terms, Brouwer's approach is an example of a strong revisionism that has the potential of both limiting and extending the actual practice. In the next section the phenomenological reasons are used to select a suitable formalism.

2.3 Implications on the formalism preference

Let us see how Brouwer's revisionist approach influenced the actual development of intuitionist formalism, and what is relevant for the current work. For the sake of

easier readability Brouwer's ideas are not presented in their original form. Instead, the modern notation is used according to Heyting's interpretation of intuitionism [58].

The familiar Boolean algebra serves as a starting point for different revisionist approaches to mathematics. Often, the easiest way of defining a new logic is through its relationship with the Boolean algebra, so the definition of Boolean logic is a natural starting point.

2.3.1 Boolean logic

Boolean logic is a system of a set A supplied with binary operators \vee , \wedge and \rightarrow , one unary operator \neg and a constant \perp . Set A with the connectives and punctuation marks forms language \mathcal{L} . Set A is then a set of variables of \mathcal{L} , $\mathbf{Var}\mathcal{L}$; variables and constants are used to build inductively defined formulae in the set $\mathbf{For}\mathcal{L}$:

- (i) \perp and $a \in \mathbf{Var}\mathcal{L}$ are formulae;
- (ii) If $a, b \in \mathbf{For}\mathcal{L}$ then $a \vee b$, $a \wedge b$, $\neg a$ and $a \rightarrow b$ are formulae too.

In classical Boolean logic, for any elements $p_0, p_1 \in \mathbf{Var}\mathcal{L}$ there are ten true propositions called the axioms of Boolean logic.

- (A1) $p_0 \rightarrow (p_1 \rightarrow p_0)$;
- (A2) $(p_0 \rightarrow (p_1 \rightarrow p_2)) \rightarrow ((p_0 \rightarrow p_1) \rightarrow (p_0 \rightarrow p_2))$;
- (A3) $p_0 \wedge p_1 \rightarrow p_0$;
- (A4) $p_0 \wedge p_1 \rightarrow p_1$;
- (A5) $p_0 \rightarrow (p_1 \rightarrow p_0 \wedge p_1)$;
- (A6) $p_0 \rightarrow p_0 \vee p_1$;
- (A7) $p_1 \rightarrow p_0 \vee p_1$;
- (A8) $(p_0 \rightarrow p_2) \rightarrow ((p_1 \rightarrow p_2) \rightarrow (p_0 \vee p_1 \rightarrow p_2))$;
- (A9) $\perp \rightarrow p_0$;
- (A10) $p_0 \vee (p_0 \rightarrow \perp)$.

The inference rules are:

Modus ponens: given formulae ϕ and $\phi \rightarrow \psi$ obtain ψ .

Substitution: given a formula ϕ obtain ϕs , where s , a substitution, is a map from $\text{Var}\mathcal{L}$ to $\text{For}\mathcal{L}$ defined inductively: $ps = s(p)$ for every $p \in \text{Var}\mathcal{L}$, $\perp s = \perp$ and $(\psi \odot \phi)s = \psi s \odot \phi s$, for $\odot \in \{\vee, \wedge, \rightarrow\}$.

Boolean logic as defined above is denoted **Cl**. It is immediately clear that **Cl** cannot be used as a reasoning framework within the realm of intuitionist mathematics: axiom (A10) is not necessarily true in intuitionist mathematics. Boolean logic is the logic of atemporal objects, the set of true formulae in the world that never changes.

2.3.2 Modal logic

A modal logic is often defined as an extension of **Cl**. However, there is a number of modal logics that are extensions of intuitionistic logics. Because most applications of modal logics to the Dempster-Shafer theory are based on the extension of classical logic, below we do not discuss intuitionistic modal logic.

A modal language \mathcal{ML} is obtained by enriching language \mathcal{L} with the new unary connective \Box and the corresponding formula formation rule.

- If ϕ is an \mathcal{ML} formula then $(\Box\phi)$ is also an \mathcal{ML} formula.

The formula formation rules for \mathcal{L} also work in \mathcal{ML} . The smallest modal logic $\mathbf{K}_{\mathcal{ML}}$ is then:

- (a) Axioms (A1)-(A10);
- (b) An additional modal axiom (A11) $\Box(p_0 \rightarrow p_1) \rightarrow (\Box p_0 \rightarrow \Box p_1)$;
- (c) The inference rules are *modus ponens*, substitution of modal formulae instead of variables and the rule of
Necessitation: given a formula ϕ , we infer $\Box\phi$.

$\mathbf{K}_{\mathcal{ML}}$ is the logic of some abstract necessity that describes the common properties characteristic of *all* interpretations of the operator \Box . It is a minimal modal logic, in a sense that any property of this logic will also be a property of any other modal logic built through defining \Box in some meaningful way. The modal language with the operator of abstract necessity is weaker than the language of a temporal logic that requires two additional operators.

There is a variety of different modal logics: temporal logic, deontic logic, epistemic logic and so on, which owe their existence to different understandings of the meaning

of the modality. Defining, interpreting and formalising modality is an amazing field which is not discussed here. Instead, a look at the semantic implications of having a modal operator is taken.

The possibility of gaining new knowledge at different states of the world is the basic accepted premise. In the logic universe gaining new knowledge equates to instantiating new variables. This possibility is best illustrated through the corresponding Kripke models. No notion of a Kripke model is introduced at the moment, so the discussion on the matter is limited to the observation that Cl can be represented by a single node and that intuitionistic and modal extensions of Cl require more complicated models. Even though K_{ML} is in some sense a minimal modal logic it is still stronger than Cl : $\text{Cl} \subset \text{K}_{\text{ML}}$ [48].

A modal extension of Cl is therefore unsuitable for building an intuitionist framework for the Dempster-Shafer theory, at least when non-constructive proofs are not allowed. It must be added that, if the authors do not reject the possibility of a non-constructive proof, different flavors of modal logic are a popular choice as shown earlier. A more detailed review of the publications on the matter is in [76].

2.3.3 Intuitionistic logic

The ‘minimal’ intuitionistic logic Int may be defined using the same propositional language \mathcal{L} as Boolean logic Cl . Int also admits axioms (A1)-(A9), but not (A10) of Cl and uses the same inference rules, *modus ponens* and substitution, as Cl . The notions of derivation and derivations from assumptions are the ones of Cl . In a superficial way, one can think of Int as of Cl without (A10).

The differences between Cl and Int run deeper than a simple exclusion of an axiom. On the formal level, excluding one axiom has a negative effect: fewer formulae are true in a weakened logic. A smaller set of true formulae is balanced by a gain in semantic. Boolean logic is a logic of atemporal objects. Intuitionist mathematics takes the epistemic aspect of the truth into account: the truth of a proposition may not be known a priori, but can be learnt later. Allowing the world to change due to *learning* new things requires a richer semantics than the one of Boolean logic. Learning new things is reflected through the concept of a possible world or a state of the world.

In terms of semantic models, Int can be defined as a set of formulae true in all possible Kripke frames with transitive nodes. Worlds may have different states at which different things are known. Hence, the same variable can be instantiated at some

worlds, but not at others. The worlds are linked through the accessibility relation in that the knowledge in the related worlds is non-contradictory. Non-contradictory knowledge means that if a variable was instantiated to some value, this value cannot be changed at a later stage. The value must stay the same in all successor nodes.

The truth of propositions is established according to intuitionist understanding:

- (i) $\phi \wedge \psi$ is true at a state (world) x if both ϕ and ψ are true at x .
- (ii) $\phi \vee \psi$ is true at a state (world) x if either ϕ or ψ is true at x .
- (iii) $\phi \rightarrow \psi$ is true at a state (world) x if for every subsequent possible state y , ϕ is true at y if and only if ψ is true at y .
- (iv) \perp is true nowhere [48].

Boolean logic **Cl** is an intuitionistic logic which consists of all formulae true at a single state of the world x . Intuitionistic propositional logic **Int** in a language \mathcal{L} is then a set of formulae that are true in all worlds and all possible configurations of such worlds. It is a well-known fact that any connected model with more than one reflexive node refutes (A10) in \mathcal{L} , see [61] and Section 3.2.2 for details.

Int is a weaker logic than **Cl**: it is known that $\text{Int} \subseteq \text{Cl}$. The properties of **Int** are quite well-known, and are not discussed here.

By now, a brief look was taken at three important logic formalisms: Boolean Logic **Cl**, Intuitionistic logic **Int** and Modal Logic $\mathbf{K}_{\mathcal{ML}}$. Among the three only the intuitionistic one does not contradict the basic premises of intuitionist mathematics. These three logics can be ordered as $\text{Int} \subseteq \text{Cl} \subseteq \mathbf{K}_{\mathcal{ML}}$. The ordering reflects only the first half of the definition of a strong revisionist approach (see page 23). The second half that mentions the potential of extension of the existing practice is realised through the superintuitionistic logic.

Logic **Int** serves as a basis for an infinite family of logics known as superintuitionistic or intermediate logics. In this work ‘superintuitionistic logic’ is used as a preferred term, partly because of the author’s personal preferences and partly to stress the fact that the logics in question are extensions of **Int**.

A superintuitionistic logic, or an si-logic for short, in language \mathcal{L} is any set L of \mathcal{L} -formulae satisfying the conditions:

- (i) $\text{Int} \subseteq L$;
- (ii) L is closed under *Modus Ponens*;

(iii) L is closed under uniform substitution.

The largest si-logic is **For** \mathcal{L} , known as inconsistent si-logic. Every si-logic that is not inconsistent is consistent. For every consistent si-logic L it is known that $\text{Int} \subseteq L \subseteq \text{Cl}$.

Reference was made to all the configurations of possible worlds while defining both $\text{K}_{\mathcal{ML}}$ and **Int**, a ‘configuration’ is represented by a Kripke model. Unlike $\text{K}_{\mathcal{ML}}$ and **Int** different si-logics are valid in different classes of models.

Up to now the discussion used a notion of *some* language \mathcal{L} defined in a fairly general way. Now, when a logic as a set of formulae is explored to a certain extent, the question of defining the connectives in \mathcal{L} can be faced.

Aside from the obvious choice of Boolean connectives, there is a whole universe of t -norm based multivalued logics. However, among the t -norm based logics only the Gödel-Dummett logic belongs to si-logics¹.

2.4 Modalities versus Beliefs

Having an additional modal operator in the propositional language is seen as an advantage by many authors. This requires defining the axioms associated with this operator that are not a part of the original Dempster-Shafer theory. Moreover, using modal extension of **Cl** requires using both reflexive and irreflexive nodes in semantic models which contradicts earlier observations about intuitionistic semantic models as it is shown below. The discussion about logics representing the Dempster-Shafer theory is mostly centred on choosing a suitable candidate from the impressive array of known modal connectives. There is no requirement for the nodes to be reflexive for modal logics.

This work uses *relational* or *possible world* semantics. In this framework relation R is the *alternativeness relation* and xRy means that y is an alternative (or possible) world for x . Under this assumption the meaning of \Box and $\Diamond = \neg\Box\neg$ on Kripke models becomes clear. $\Box\phi$ is true at a node w if ϕ is true at all nodes reachable from w , $\Diamond\phi$ is true if ϕ is true in at least one node reachable from w . Given this, the attention to whether a node is reflexive or not should be paid.

Consider now the simplest single-node models in Figure 2.1. In the picture reflexive nodes are empty circles and irreflexive ones are filled. The model in both cases consists of a single node w . The formulae true at the node are listed on the left

¹The author is very grateful to C.Fermüller for drawing his attention to this fact

of it, and the ones that are false on the right.

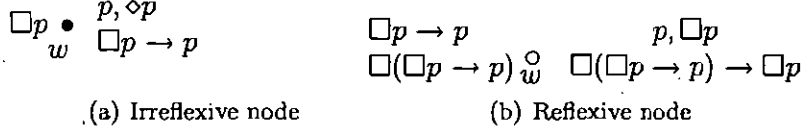


Figure 2.1: A simple modal model

For both frames, $\mathfrak{F} = \langle W, R \rangle$ with $W = \{w\}$ and $V(p) = \emptyset$. Relations R are different for different models: for the model in Figure 2.1(a) $R = \emptyset$ and in Figure 2.1(b) $R = \{(w, w)\}$. This ‘minor’ difference leads to a significant semantic difference between the models. The necessity operator is validated on an irreflexive node, but is refuted on a reflexive one. The sequence can be continued. A few formulae are listed in both cases.

Consider now an irreflexive node in the Dempster-Shafer theory context. Even if all the belief is attributed to the node such that p is false, p still must be true somewhere else.

2.5 The formalism choice

The question of the best logical formalism for interpreting the Dempster-Shafer theory is likely to stay open for a long time, mostly because several different viewpoints have resulted in feasible results. The choice of a particular formalism is still largely determined by the factors outside of the Dempster-Shafer theory proper. Often such a choice is based on focusing attention on some particular aspect of the theory. The choice is based on some attempt to interpret the beliefs with the aid of either modalities or the truth values of propositions. This approach is ‘dangerous’ because of the fundamental difference between the two concepts. As Hajék puts it [76]:

Truth degrees in fuzzy logic must be clearly distinguished from belief degrees in the Dempster-Shafer theory.

Fuzzy logic is the logic of comparative truths that are understood as truth-functional. Belief degrees are not truth-functional.

The statement above does not explicitly mean that there is no connection between the degrees of truth and the degrees of belief. An si-logic can be used to find a degree of

truth of a proposition with a belief attributed to it. However, to make the exposition clearer, the notions of modality and belief are kept separate in this work.

The author argues in favour of taking a more general approach. The claim is substantiated through an attempt to understand the nature of mathematical objects that constitute the Dempster-Shafer theory universe from a phenomenological point of view first. The presented argument is by no means exhaustive. It rather shows a fairly obvious distinction which, if noticed early enough, leads a researcher in a different direction. The approach yields a practical result: the models induced according to the principles described above validate **Int**. The rest of the work presents the approach in detail.

Chapter 3

Semantic models for the belief theory

This chapter describes the semantic representation of Dempster-Shafer frames of discernment and demonstrate how to calculate the beliefs in different propositions from these models. The given procedure also provides a meaningful interpretation of the Dempster-Shafer evidence combination rule.

The chapter is organised as follows. The first two sections provide the necessary background to the Dempster-Shafer theory and the semantic models. Several relevant examples that illustrate the features of the theories in question are analysed. Once the necessary background is introduced, the formal description of the proposed procedure is developed and illustrated on a set of examples. After demonstrating that the beliefs induced on the semantic models are the same as in the original setup the author progresses to the formal verification of the approach.

3.1 Revisiting the theory of beliefs

Two theories have inspired this research: the Dempster-Shafer theory and the theory of Semantic models also known as Kripke models. The review begins with introduction to the Dempster-Shafer theory. The author tried to avoid repeating the information already presented in the earlier chapters as much as possible, but some concepts mentioned casually in the earlier chapters need the formal definitions provided in this section.

3.1.1 The formalism

The exposition below is very basic, and the quoted results can be found in any book on the topic. To minimise the possible distortions of the original concept, the definitions are quoted according to Shafer's essay [9] unless stated otherwise.

The Dempster-Shafer theory initially was built to deal with finite sets [9, 13, 14]. Later, the infinite universe was introduced by Shafer himself in his doctoral thesis [1]. From the practical viewpoint there is still nothing wrong with starting with the finite sets.

The Dempster-Shafer theory started as a generalisation of Bayesian theory and was defined through amending certain parts. Both Bayesian and the Dempster-Shafer theory describe their universe through set Θ : a finite non-empty set that, along with all of its subsets, in the Dempster-Shafer theory is called the *frame of discernment*. Guan and Bell in [96] suggest to think about Θ as of '*the set of all possible true values that a quantity we are interested in can take.*'

Two fundamental concepts of Bayesian statistics are the Bayesian (probabilistic) density and the Bayesian function. A function $d : \Theta \rightarrow [0, 1]$ is a Bayesian density if $\sum_{x \in \Theta} d(x) = 1$.

A function $bay : 2^\Theta \rightarrow [0, 1]$ is a Bayesian function if the following three conditions are met:

1. $bay(\emptyset) = 0$;
2. $bay(\Theta) = 1$;
3. $bay(A \cup B) = bay(A) + bay(B)$ whenever $A \cap B = \emptyset$.

Condition 3 can be easily generalised to finite unions of any subsets:

3a.

$$bay(A_1 \cup \dots \cup A_n) = \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} bay(\cap_{i \in I} A_i). \quad (3.1)$$

In the case of two overlapping sets, formula (3.1) becomes a more familiar $bay(A \cup B) = bay(A) + bay(B) - bay(A \cap B)$. The Bayesian functions are used to describe probabilities of events in a probabilistic space. There are several equivalent alternative ways to state the last condition. Bayesian density functions and Bayesian functions are in a one-to-one correspondence. The Dempster-Shafer theory operates with the *belief functions* defined on page 17.

There is an obvious parallel between Bayesian functions and beliefs: in a two-set case condition 3 from Definition 2.1.1 becomes either $Bel(A \cup B) \geq Bel(A) +$

$Bel(B) - Bel(A \cap B)$ or $Bel(A \cup B) \geq Bel(A) + Bel(B)$ in case $A \cap B = \emptyset$. All the Bayesian functions are belief functions but not vice-versa, as the Dempster-Shafer theory generalises the Bayesian statistics. Belief functions are too general to be immediately applicable, while basic probability assignments are more manageable and allow one to calculate corresponding beliefs. Basic probability assignments defined on page 17 are neither additive nor monotone.

Equation (2.1) is used as an alternative definition of a belief function. The definitions demonstrate the dual nature of basic probability assignments and belief functions. The mathematical apparatus dealing with this duality allows for conversion in either direction. On the other hand, when an actual frame of discernment Θ is considered, the problem of either finding a belief function or assigning basic probabilities to different elements of 2^Θ is far from obvious.

Depending on the need, other 'belief-like' functions are used: commonality functions, plausibility functions, doubt functions and ignorance, a detailed review of which can be found in [96]. Given a mass assignment, the rest of the arsenal can be developed easily. Of those only the plausibility function is used in the current work:

$$pls(A) = 1 - bel(\bar{A}), \quad (3.2)$$

where A is a subset of Θ , and $\bar{A} = \Theta \setminus A$ is a set complement of A .

The Bayesian rule updates known probabilities of events that are not independent.

Definition 3.1.1 (Bayesian Rule) *Given a Bayesian function bay and $\emptyset \neq B \subseteq \Theta$ one can calculate the conditional probability $bay(X | B)$ of $X \in 2^\Theta$ under B is $bay(\cdot | B) : 2^\Theta \rightarrow [0, 1]$ and for $bay(B) > 0$ the Bayesian rule applies:*

$$bay(X | B) = \frac{bay(X \cap B)}{bay(B)}.$$

Dempster-Shafer evidence combination rule combines known beliefs in a more relaxed setting of the Dempster-Shafer theory:

Definition 3.1.2 (Dempster-Shafer evidence combination rule) *Let m_1 and m_2 be basic probability assignments on the same frame Θ . Suppose*

$$E = \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) \leq 1. \quad (3.3)$$

Denote

$$N = \sum_{X \cap Y \neq \emptyset} m_1(X)m_2(Y). \quad (3.4)$$

Then the function $m : 2^\Theta \rightarrow [0, 1]$ defined by

$$m(\emptyset) = 0, \quad (3.5)$$

and

$$m(A) = \frac{1}{N} \sum_{X \cap Y = A} m_1(X)m_2(Y) \quad (3.6)$$

for all subsets $A \neq \emptyset$ of Θ is a basic probability assignment.

Assignment m is called the orthogonal sum of m_1 and m_2 and is denoted $m = m_1 \oplus m_2$.

The theorem below gives a formal interpretation of the analogy between the Bayesian and Dempster-Shafer rules [96].

Theorem 3.1.1 *Let m_B be a mass function such that*

$$m_B(B) = 1, \quad m_B(\text{elsewhere}) = 0$$

Then

1. *A mass function m and m_B are combinable iff*

$$bel(\overline{B}) < 1.$$

2. *If m and m_B are combinable, denote*

$$bel(X|B) = (bel \oplus bel_B)(X), \quad pls(X|B) = (pls \oplus pls_B)(X).$$

Then

$$bel(X|B) = \frac{bel(X \cup \overline{B}) - bel(\overline{B})}{1 - bel(\overline{B})} = \frac{pls(X \cap B)}{pls(B)}$$

for all $B \subseteq \Theta$.

3.1.2 Calculating the beliefs

The problem is addressed in several steps. In the beginning, the simplest situation with the evidence in support of only one statement is taken, as the analysis evolved

more complicated situations are introduced. The belief function in the simplest case needs to reflect only how much belief is attributed exactly to the only proposition and make sure that the total belief over the whole of the frame of discernment is equal to one. Belief functions of this type are called *simple support functions*. The functions obtained by combining different simple support functions are called *separable support functions*. Separable support function form a subset of a more general class of support functions, which are a particular case of belief functions. The distinction between different types of belief functions is not very important at the moment. The relation between different types of belief functions is given in Figure 3.1. The simplest support

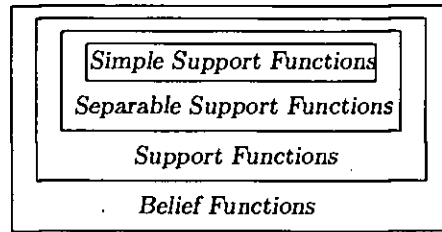


Figure 3.1: Support functions

function that describes the beliefs attributed to a specific subset A of Θ requires the following basic probability assignments:

$$m(A) = s_1, \quad m(\Theta) = 1 - s_1, \quad m(B) = 0, \quad (3.7)$$

whenever $B \neq A$. It is important to note that $m(\neg A) = 0$, basic probability assignments are not probabilities. The analog of probabilities in this case is the support function. As a simple support function is a belief function, the probability assignments give

$$S(B) = \begin{cases} 0 & \text{if } B \neq A, B \neq \Theta; \\ s_1 & \text{if } B = A; \\ 1 & \text{if } B = \Theta. \end{cases}$$

This simple support function is said to be centred on B .

As analysis shows, a large proportion of possible situations can be modelled using simple and separable support functions. However, an interpretation that is only capable of representing the frames of discernment with separable support functions is too restrictive. This restriction should be avoided if possible. The representational limits of the approach are discussed in some detail in Chapter 4.

The background introduced to the point allows one to see how the beliefs in

different propositions can be calculated, and how they can be updated with the aid of the Dempster-Shafer evidence combination rule. The application of the Dempster-Shafer rule is demonstrated on the same humorous examples that Shafer used in his book [9], and that later became quite popular among the wider research community. This work follows the tradition and employs the familiar stories to illustrate both known facts about the theory and to show that the proposed procedures are adequate to the setting.

In his book Shafer used four short stories: the burglary of the sweetshop, the cabbage seed, the alibi and the biased coin. As his analysis progressed, he showed that the alibi and the biased coin were effectively the same setting, so only the first three cases are used in this work. The stories and the calculations are according to [96].

The burglary of the sweetshop

Sherlock Holmes investigates the burglary of the sweetshop. Initial evidence shows that the burglar is left-handed. Sherlock Holmes attributes a degree of belief s_1 to the fact. Later, new evidence emerges: it was an insider job. The degree of confidence about this new piece of evidence is s_2 . There is a left-handed clerk in the shop who comes under suspicion. What is the degree of belief in this clerk's guilt?

The burglary of the sweetshop example involves two separate, but not contradictory pieces of evidence. There are two separate frames of discernment to be combined. The first frame is $\Theta_1 = \{L, \neg L\}$ corresponding to the evidence about the burglar being left-handed or otherwise. Similarly, $\Theta_2 = \{I, \neg I\}$ reflects the evidence that the burglar may or may not be an insider. To calculate the belief in the left-handed clerk's guilt, the basic probability assignments on each of the frames of discernment are needed first.

It is already known that $m_1(\{L\}) = s_1$ meaning that $m_1(\{\Theta_1\}) = 1 - s_1$. Similarly, on the second frame $m_2(\{I\}) = s_2$ and $m_2(\{\Theta_2\}) = 1 - s_2$. The frame of discernment Θ that is produced after combining Θ_1 and Θ_2 is $\Theta = \{L \wedge I, L \wedge \neg I, \neg L \wedge I, \neg L \wedge \neg I\}$, while the probability assignments for the elements Θ are found according to the procedure given in Definition 3.1.2. From now on this frame is referred to as $\Theta = \{LI, LO, RI, RO\}$ making use of the obvious: $L \wedge I$ is left-handed insider, LI ; $L \wedge \neg I$ is left-handed non-insider or left-handed outsider, LO etc. To make things easier, the orthogonal sum of probability assignments is calculated with help of the intersection Table 3.1. The cells in the body of the table correspond to the intersections of the elements in the frame of discernment Θ produced by combining Θ_1 and Θ_2 . Following

Table 3.1: Intersection table for burglary of sweetshop

$m_1 \oplus m_2$	$\{LI, RI\}s_2$	$\Theta 1 - s_2$
$\{LI, LO\}s_1$	$\{LI\}s_1s_2$	$\{LI, LO\}s_1(1 - s_2)$
$\Theta 1 - s_1$	$\{LI, RI\}(1 - s_1)s_2$	$\Theta(1 - s_1)(1 - s_2)$

the tradition, rows and columns are labelled with both sets and their probability mass assignments.

Reading the results from Table 3.1 gives

$$(m_1 \oplus m_2)(\{LI\}) = s_1s_2, (m_1 \oplus m_2)(\{LI, LO\}) = s_1(1 - s_2),$$

$$(m_1 \oplus m_2)(\{LI, RI\}) = (1 - s_1)s_2, (m_1 \oplus m_2)(\Theta) = s_1(1 - s_2),$$

$$(m_1 \oplus m_2)(elsewhere) = 0.$$

Knowing the probability mass assignments on the new frame allows one to calculate the belief, plausibility and any other function of interest.

In case of the left-handed clerk, the confidence in his guilt is estimated by looking at the interval between the value of the belief and the plausibility functions defined in equations (2.1) and (3.2).

$$bel_{m_1 \oplus m_2}(\{LI\}) = s_1s_2,$$

$$pls_{m_1 \oplus m_2}(\{LI\}) = 1 - bel_{m_1 \oplus m_2}(\Theta \setminus \{LI\}) = 1 - bel_{m_1 \oplus m_2}(\{LI, LO, RO\}) = 1 - 0 = 1,$$

thus placing the confidence values into the interval

$$[bel_{m_1 \oplus m_2}(\{LI\}), pls_{m_1 \oplus m_2}(\{LI\})] = [s_1s_2, 1].$$

The cabbage seed

A mathematician of questionable gardening skills plants a seed in the pot. When first shoots are sprouted, the gardener puts support of c_1 that it is a cabbage (statement A). On a closer inspection the plant has two leaves, so it is a member of *brassica* genus with support of c_2 (this statement is B). The frame of discernment Θ in this case is the set of all plants. To simplify the notation put $A = \{cabbages\}$ and $B = \{brassicas\}$. Common sense also suggests that $c_2 > c_1$, even though this condition is not formally required.

Table 3.2: Probability mass assignments for the cabbage seed

$m_1 \oplus m_2$	$B \ c_2$	$\Theta \ 1 - c_2$
$A \ c_1$	$A \ c_1 c_2$	$A \ c_1(1 - c_2)$
$\Theta \ 1 - c_1$	$B \ (1 - c_1)c_2$	$\Theta \ (1 - c_1)(1 - c_2)$

This example is different from the burglary of the sweetshop. The two pieces of evidence still do not contradict each other, but one piece *includes* another. If the plant is a cabbage, then it must belong to brassica genus, but not vise-versa.

Combining the evidence is then finding the orthogonal sum of the following mass assignments

$$m_1(A) = c_1, \ m_1(\Theta) = 1 - c_1, \ m_1(\text{elsewhere}) = 0$$

and

$$m_2(B) = c_2, \ m_2(\Theta) = 1 - c_2, \ m_2(\text{elsewhere}) = 0.$$

The intersection table for m_1 and m_2 is given in Table 3.2 and it follows that

$$(m_1 \oplus m_2)(A) = c_1 c_2 + c_1(1 - c_2) = c_1; \ (m_1 \oplus m_2)(B) = (1 - c_1)c_2,$$

$$(m_1 \oplus m_2)(\Theta) = (1 - c_1)(1 - c_2), \ (m_1 \oplus m_2)(\text{elsewhere}) = 0.$$

Calculating probability intervals for both A and B then gives

$$\text{bel}_{m_1 \oplus m_2}(A) = (m_1 \oplus m_2)(A) = c_1,$$

$$\text{bel}_{m_1 \oplus m_2}(B) = (m_1 \oplus m_2)(A) + (m_1 \oplus m_2)(B) = c_1 + (1 - c_1)c_2 = c_1 + c_2 - c_1 c_2$$

and, similarly,

$$\text{pls}_{m_1 \oplus m_2}(A) = 1 - \text{pls}_{m_1 \oplus m_2}(\Theta \setminus A) = 1 - 0 = 1;$$

$$\text{pls}_{m_1 \oplus m_2}(B) = 1 - \text{pls}_{m_1 \oplus m_2}(\Theta \setminus B) = 1 - 0 = 1.$$

The confidence intervals are then

$$[\text{bel}_{m_1 \oplus m_2}(A), \text{pls}_{m_1 \oplus m_2}(A)] = [c_1, 1],$$

$$[\text{bel}_{m_1 \oplus m_2}(B), \text{pls}_{m_1 \oplus m_2}(B)] = [c_1 + c_2 - c_1 c_2, 1].$$

The belief intervals above show that the new evidence did not change anything about

our beliefs in the fact that the plant could be a cabbage: they stayed the same. The belief in brassica was strengthened after the evidence about the plant being a cabbage was incorporated. To sum up, the beliefs in the detailed proposition did not change, but the belief in the general statement increased.

Alibi

The last example demonstrates the third possibility: two contradictory pieces of evidence. In terms of sets, such situation corresponds to two pieces of evidence represented by two non-overlapping sets.

One person is arrested on a murder charge. A reliable witness provides an alibi for the accused with a degree of support s_1 for the claim of innocence. On the other hand, circumstantial evidence gives s_2 support for the accused's guilt. This evidence can be represented as two frames $\Theta_1 = \{I, \neg I\}$, where I is the set of all people who have an alibi, and $\Theta_2 = \{G, \neg G\}$, where G is the set of all people who are guilty according to circumstantial evidence. The resulting frame Θ incorporates both Θ_1 and Θ_2 . The new frame will have two different probability mass assignments: m_1 for the first piece of evidence and m_2 for the second. Two pieces of evidence, however, contradict each other. In terms of sets within Θ_1 and Θ_2 it means that there are sets whose intersection is empty. Indeed, consider $I \cap G$. A person can either be guilty or have an alibi, but not both, so $I \cap G = \emptyset$. Both $m_1(I)$ and $m_2(G)$ are not zeros, so some of the earlier beliefs are attributed to the empty set or just wasted. The probability assignments are given by

$$m_1(\{I\}) = s_1, m_1(\{\Theta_1\}) = 1 - s_1, m_1(\text{elsewhere}) = 0,$$

$$m_2(\{G\}) = s_2, m_2(\{\Theta_2\}) = 1 - s_2, m_2(\text{elsewhere}) = 0.$$

Table 3.3 is the intersection table for this case. One of the cells contains the empty set, but the belief attributed to this cell is not zero. This wasted belief must be balanced by introducing the normalising coefficient N , which in the earlier examples was just one. For alibi example, N is calculated according to $N = 1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) = 1 - m_1(I)m_2(G) = 1 - s_1s_2$. To get the actual beliefs in different statements in the combined frame of discernment, the entries in the intersection table must be divided by N .

Table 3.3: Intersection table for alibi

$m_1 \oplus m_2$	$\{G\} \ s_2$	$\Theta \ 1 - s_2$
$\{I\} \ s_1$	$\emptyset \ s_1 s_2$	$\{I\} \ s_1(1 - s_2)$
$\Theta \ 1 - s_1$	$\{G\} \ (1 - s_1)s_2$	$\Theta \ (1 - s_1)(1 - s_2)$

The normalised probability assignments then become

$$(m_1 \oplus m_2)(\{I\}) = \frac{s_1(1 - s_2)}{1 - s_1 s_2},$$

$$(m_1 \oplus m_2)(\{G\}) = \frac{(1 - s_1)s_2}{1 - s_1 s_2},$$

$$(m_1 \oplus m_2)(\{\Theta\}) = \frac{(1 - s_1)(1 - s_2)}{1 - s_1 s_2},$$

and zero elsewhere.

The belief in the accused's innocence

$$bel_{m_1 \oplus m_2}(\{I\}) = (m_1 \oplus m_2)(\{I\}) = \frac{s_1(1 - s_2)}{1 - s_1 s_2},$$

while the belief in his guilt is

$$bel_{m_1 \oplus m_2}(\{G\}) = (m_1 \oplus m_2)(\{G\}) = \frac{(1 - s_1)s_2}{1 - s_1 s_2}.$$

The plausibility of either outcome

$$pls_{m_1 \oplus m_2}(\{I\}) = 1 - bel_{m_1 \oplus m_2}(\Theta \setminus \{I\}) = 1 - bel_{m_1 \oplus m_2}(\{G\}) = \frac{1 - s_2}{1 - s_1 s_2}$$

and

$$pls_{m_1 \oplus m_2}(\{G\}) = 1 - bel_{m_1 \oplus m_2}(\Theta \setminus \{G\}) = 1 - bel_{m_1 \oplus m_2}(\{I\}) = \frac{1 - s_1}{1 - s_1 s_2}.$$

The probability intervals for each possibility are

$$[bel_{m_1 \oplus m_2}(\{I\}), pls_{m_1 \oplus m_2}(\{I\})] = \left[\frac{s_1(1 - s_2)}{1 - s_1 s_2}, \frac{1 - s_2}{1 - s_1 s_2} \right],$$

and

$$\{bel_{m_1 \oplus m_2}(\{G\}), pls_{m_1 \oplus m_2}(\{G\})\} = \left[\frac{s_2(1-s_1)}{1-s_1s_2}, \frac{1-s_1}{1-s_1s_2} \right].$$

The outcome of evidence combination is quite predictable, beliefs in both possibilities are perturbed, and each side's case is weaker than it was in the beginning:

$$s_1 > \frac{s_1(1-s_2)}{1-s_1s_2} \quad \text{and} \quad s_2 > \frac{s_2(1-s_1)}{1-s_1s_2}.$$

3.2 Semantic models

3.2.1 Models and validity of formulae

A basic definition of an intuitionistic Kripke model was first given in Chapter 2. The present discussion demands more background knowledge, so some time will be devoted to filling in the lacunae. The Kripke model is a convenient tool that is used for different purposes and in different settings. Since their introduction, Kripke models have earned a fully deserved popularity among the researchers in philosophy, mathematics and computer science. Given the variety of applications, the definitions of Kripke models differ slightly from researcher to researcher. An interesting overview of various definitions of the Kripke model is given in [85]. A detailed analysis of Kripke's philosophy can be found either in his own works, say [69], or in monograph [70].

Definition 2.1.3 is closest to the one in [48]. Kripke models provide understanding of the relational semantics associated with the Dempster-Shafer theory. Using Kripke models as a starting point has been a popular approach since the 1970's in many different settings. A good review of possible directions of the enquiry and basic results can be found in [97].

The background given in the two previous chapters meant to help the reader to develop some kind of intuition about the building blocks of the Dempster-Shafer theory. A few observations below are intended to help develop a similar feel for the semantic models. Recall that a Kripke model is triple $\mathfrak{F} = \langle W, R, V \rangle$, where W is a set of possible worlds, R is a partial order (accessibility relation) on W , and V is a valuation that maps statements from \mathcal{L} into $\text{Up}W$.

To get a better intuition about the nature of relation R consider a local newsagent. Let W be the set of all newspapers on sale, \mathcal{L} be the set of reported news. Valuation $V(p)$ shows which newspapers reported news $p \in \mathcal{L}$; relation R is quoting: xRy means that newspaper y , perhaps a small regional tabloid that cannot afford to hire many

reporters, can quote newspaper x , which can be a national broadsheet that has a lot of correspondents and bureaus around the World.

A Kripke frame is a Kripke model without a valuation function: a Kripke frame \mathfrak{F} is a tuple $\mathfrak{F} = \langle W, R \rangle$, where W and R are as in Definition 2.1.3. Alternatively, one can refer to a model as a pair $\langle \mathfrak{F}, V \rangle$.

The relationship between the propositions of \mathcal{L} and their semantic model counterparts is straightforward. A proposition can be instantiated at a node, and then it stays true in all of the node's successors. The relationship between the truth of statements in the real world is somehow more complicated. Modelling the reality using Kripke models as defined above leads to minor discrepancies of a particular type. The statement '*The solar system has eight planets*' was known to be false until recently, because Pluto was classified as a planet and the count was nine. An example of the converse can be constructed easily too: '*Anatoli Karpov is the world chess champion.*' It is important though to remember that the model in question does not attempt to describe the world we live in, so the latter is impossible. In other words, if one built a Kripke model that describes the ranking of the chess players any time between 1975 and 1985 (the time when Karpov was the undisputed World champion), he would not be able to adjust it to the modern reality.

The limitation outlined in the previous paragraph is not too important for the purposes of evidence combination, though. The author is more interested in combining old and newly learnt evidence rather than tracing the changing values of different statements. The impossibility of changing the truth assignment of some variables can be easily overcome by introducing an extra piece of evidence that contradicts the evidence introduced earlier. This situation was analysed when the evidence was combined for the alibi on page 39.

Kripke models are capable of representing richer semantics than the one of Cl, which is the set of formulae true at a single node. The gain in representational power is accompanied with increased complexity because it requires a special valuation function that keeps track of the formulae valid at different states of the world. The valuation function also makes checking a validity of a formula — a much more complicated business than in the case of Cl. Developing an inferential apparatus requires the terminology that describes the truth and validity of formulae. The first concept to be introduced is the inductively defined relation \models .

Definition 3.2.1 (Validity of formulae) Let $\mathfrak{M} = \langle W, R, V \rangle$ be a Kripke model, let $x \in W$, $\phi, \psi \in \text{For } \mathcal{L}$, $p \in \text{Var } \mathcal{L}$, then $(\mathfrak{M}, x) \models \chi$, which is read ' χ is true at x '

in \mathfrak{M} if:

$$\begin{aligned}
(\mathfrak{M}, x) \models p & \quad \text{iff} \quad x \in V(p); \\
(\mathfrak{M}, x) \models \phi \wedge \psi & \quad \text{iff} \quad (\mathfrak{M}, x) \models \phi \text{ and } (\mathfrak{M}, x) \models \psi; \\
(\mathfrak{M}, x) \models \phi \vee \psi & \quad \text{iff} \quad (\mathfrak{M}, x) \models \phi \text{ or } (\mathfrak{M}, x) \models \psi; \\
(\mathfrak{M}, x) \models \phi \rightarrow \psi & \quad \text{iff} \quad \text{for all } y \in W \text{ such that } xRy, (\mathfrak{M}, y) \models \phi \text{ implies } (\mathfrak{M}, y) \models \psi; \\
(\mathfrak{M}, x) \not\models \perp.
\end{aligned}$$

There are several degrees of a formula's validity within the realm of intuitionistic or si-logic. The weakest degree of validity is the truth at a point in a model. In this case it is said that a formula is *satisfied* in a model. Following [48], we denote by $V(\phi)$ the *truth set* of a formula

$$V(\phi) = \{x \in W \mid (\mathfrak{M}, x) \models \phi\}.$$

Thus, $V(\phi)$ is then the set of all nodes where the formula is valid.

A formula ϕ is *true* in a model if $(\mathfrak{M}, x) \models \phi$ for every $x \in W$. The validity in a frame is even stronger. A formula ϕ is *satisfied* in a frame \mathfrak{F} if it is $(\mathfrak{M}, x) \models \phi$ for some x and some \mathfrak{M} , such that $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$. A formula is true at a point $x \in \mathfrak{F}$ if $(\mathfrak{M}, x) \models \phi$ for all models such that $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$. Finally, a formula ϕ is *valid* in a frame \mathfrak{F} if it is true in all models based on \mathfrak{F} . A formula ϕ valid in a frame \mathfrak{F} is denoted $\mathfrak{F} \models \phi$.

The notion of refutation is symmetric to the notion of truth: a formula ϕ is refuted in a model \mathfrak{M} if it is not true in it. \mathfrak{M} is then a counter model to ϕ . Similarly, a formula is refuted in a frame \mathfrak{F} if it is not true in it.

To make manipulating models easier they are often represented graphically: the points are represented by circles, the relation R by arrows, the propositions true at a point are listed below it to the right of the vertical line, the ones that are false — to the left. The reflexive pairs xRx are not listed. For example, model $\mathfrak{M} = \langle W, R, V \rangle$, where $W = \{w_1, w_2\}$, $\mathcal{L} = \{p, q, \vee\}$, $R \ni (w_1, w_2)$, $V(p) = \{w_1, w_2\}$ and $V(q) = \{w_2\}$ is in Figure 3.2.

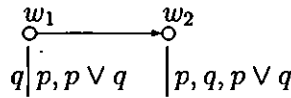


Figure 3.2: A simple Kripke model

3.2.2 Refuting *tertium non datur*

The notion of a counter model allows one to demonstrate how the objectives stated in Section 2.3.3 are met by a logic represented by Kripke frames with more than one node. The example below is famous and used by many researchers. In this work it is quoted according to [61]. Consider a model such that $W = \{w_0, w_1\}$, and the relation

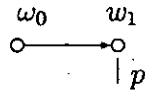


Figure 3.3: The model refuting $p \vee (p \rightarrow \perp)$.

$R = (\{w_0, w_1\})$. Let $V(p) = \{w_1\}$ and let $V(\cdot) = \emptyset$ for any other propositional letter. Graphically, this model is just a pair of nodes joined by a single edge given in Figure 3.3. Let us now compute the truth set $V(p \vee (p \rightarrow \perp))$. From Definition 2.1.3 it follows that

$$V(p \vee (p \rightarrow \perp)) = V(p) \cup V(p \rightarrow \perp) = \{w_1\} \cup V(p \rightarrow \perp) = \{w_1\} \cup \emptyset = \{w_1\}.$$

But this means that $p \vee (p \rightarrow \perp)$ is not valid at w_0 .

Formula $p \vee (p \rightarrow \perp)$ is refuted at the *root* of a simple model. In this context, the root is the node that does not have any predecessors. Refuting the formula at a root of a model means that, as long as a model contains at least two nodes and a variable instantiated to being true at the root's immediate successor, it can serve as a counter model for $p \vee (p \rightarrow \perp)$. The only models that do not refute *tertium non datur* are the models that consist of a single node or of a collection of isolated single nodes. Such models are not expressive enough for our purposes. Later, it will be shown that models that are collections of isolated nodes correspond to the situations that precede evidence combination and thus are only a subset of the models that interest us.

3.3 Translating frames of discernment to Kripke models

The procedure developed in this section should unambiguously and meaningfully link the Dempster-Shafer theory and Kripke models. Unambiguous link means that the correspondence between Kripke models and frames of discernment should be a one to

one and onto. Meaningful, in the context, means that one should be able to calculate the beliefs using both the frame of discernment and its corresponding Kripke model without any loss of information.

The construction begins with defining the Kripke model building procedure based on the knowledge of the frame of discernment. Then, the examples from Section 3.1 are presented by Kripke models and the belief functions induced on the frames are calculated. Finally, it is proved that the belief functions induced on Kripke models are always equal to the ones calculated directly from the frames of discernment, thus demonstrating that the proposed procedure gives a meaningful semantic interpretation of the Dempster-Shafer theory. The results presented in this section were first published in [84].

3.3.1 Constructing set W

Kripke models for frames of discernment are built using an intuitive procedure. The presentation starts with a very general idea about what is wanted from the semantic interpretation of frames of discernment and show how the objectives may be achieved formally. First of all, the semantic models must serve as inferential tools that facilitate calculating the support of statements known to be true. The models should also help to validate or refute formulae whose support is not given explicitly, but could be inferred from the known premises.

It is already known that attributing some belief to a particular statement should at least amount to instantiating a variable in a Kripke model. Cl is the set of formulae that is validated at a single node Kripke model. If there is only support for a single statement, and the statement is certain, the situation must be in the realm of Cl . Whenever more than one statement is supported, the semantics is richer, and the departure from Cl may be needed. The obvious way to represent such situation with semantic models is to have different nodes for different statements that are supported.

The next objective is to represent evidence combination and update. Introducing new evidence affects relation R and valuation V . Relation R should show possible paths of consistent reasoning, while the valuation function is responsible for determining at which nodes a particular formula is valid. The degree of belief in each proposition should be determined by the total belief assigned to the nodes where it is true. First, one needs to decide which nodes to include and then analyse their relationships, thus constructing R and V .

The procedure is twofold: it represents mass assignments on some existing frames

of discernment, and it provides a tool for combination of evidence represented by two different mass assignments. While translating frames of discernment to elementary Kripke models is not too exciting, the evidence combination in terms of Kripke models is quite effective and leads to constructing an inferential apparatus different from CI.

In the Dempster-Shafer theory universe there is a frame of discernment Θ , which is a set along with some portion of its power subset 2^Θ discerned by the particular setup. The mass assignments over elements of 2^Θ form the support function S . The mass assignments also allow one to calculate the beliefs for different propositions, but recall the remark about the general belief functions: they are a very broad class and always useful from practical viewpoint. The support functions are more manageable while being general enough to describe almost any conceivable setup. It is important to remember that different support functions give rise to different Kripke models. In some sense, the purpose of the procedure is not to model a frame of discernment itself, but to find the best representation of support functions defined on frames.

Representing support functions requires amending the definition of the objects that are studied. Instead of operating with semantic models the focus of attention is now switched to the semantic models with mass assignments. It will be demonstrated shortly that the structure of a model depends on the support function that induces it. The starting point in this case is the frame of discernment and the support function over it.

Assigning probability masses to different subsets of Θ is unambiguously described through the propositions in some language \mathcal{L} , which serves as a bridge between the Dempster-Shafer theory universe and the universe described by the Kripke models.

As defined earlier, a Kripke model is a triple $\mathfrak{F} = \langle W, R, V \rangle$, where W is the set of possible worlds, each with statements in \mathcal{L} that are true in it. Relation R shows the possible direction of inference, while valuation V links $\text{Var}\mathcal{L}$ and $\text{Up}W$. The Kripke models for the Dempster-Shafer theory includes information about the mass assignments for each possible world. In order to distinguish between the Kripke models as a general concept and the Kripke models analysed in this work, a notion of Kripke model with mass assignments is used.

Definition 3.3.1 (Kripke model with mass assignments) *A Kripke model with mass assignments is a four tuple $\mathfrak{F}_m = \langle \mathfrak{F}, m_w \rangle = \langle W, R, V, m_w \rangle$, where W , R , and V are elements of an intuitionistic Kripke model defined in Chapter 2, and $m_w : W \rightarrow [0, 1]$ is the mass assignment on the elements of set W .*

To get a meaningful transition from Θ to \mathfrak{F}_m one needs to form set W and then

determine the relationships among its elements, thus constructing R . Valuation V is partially determined by probability mass assignments.

Set inclusion affects the beliefs: assigning some probability mass to a subset of a set changes (increases) the set's support. The relationship does not affect the subsets — adding beliefs to more specific propositions increases belief in more general ones. In terms of semantic models it means that the nodes that represent more general propositions must see the nodes that represent more specific ones. The last observation gives us a hint about the way to construct relationship R .

Recall, that the core of a support function S on frame of discernment Θ is a collection of elements of 2^Θ such that their mass assignments are not zero: $X \in \mathcal{C}(S)$ if $X \in 2^\Theta; m(X) > 0$. One can think of core elements of a frame of discernment as of statements whose validity is known. It is then natural to expect every core element to translate into a separate node with some probability mass assigned to it. It will soon be shown that the intuition is correct, but requires some formal trickery in order to make it agree with the rest of requirements.

The first element in the pair $\langle \mathfrak{F}, m_w \rangle$ from definition 3.3.1 is then determined by the core of the support function represented through the node mass assignments m_w . In other words, the support functions with identical cores, but not necessarily identical mass assignments of the elements, result in the same model \mathfrak{F} . In the discussion that follows the abbreviated version of definition 3.3.1 will be used. Whenever a reference made to a Kripke model $\mathfrak{F} = \langle W, R, V \rangle$ it includes all possible mass assignments m_w that represent support function S such that its core $\mathcal{C}(S)$ induces set W . The abbreviation makes sense, because the models that represent support functions with the same core are semantically equivalent: they validate the same set of formulae, even though different beliefs can be attributed to the same statements depending on m_w .

Thus, to translate support function S to the language of Kripke models one has to assign a separate node $w \in W$ to each element of $\mathcal{C}(S)$. On the other hand, even the simplest support functions often assign some probability mass to the *whole* of the frame of discernment Θ . While perfectly reasonable from the theory of evidence viewpoint, such an assignment is not very convenient from the point the semantic models' point of view — it leads to existence of a 'supernode' whose support is one and thus equals to the support of the model's totality. In semantic terms it is more convenient to have a node where none of the core elements is known to be true. This node is different from the node where all the core elements are known to be false or the node that represents the whole of Θ . The meaning of this node will be illustrated

in more detail once all the necessary definitions are given.

The requirement to assign a node to each core element of S formally amounts to a non-empty valuation for each such element that includes a node that validates this element and nothing else, meaning that for any $X \in C(S)$, there is a node $w \in W$ such that $w \in V(r)$ and $w \notin V(t)$, where propositions r and t are $r = "x \in X"$ and $t = "x \in Y"$ for any set $Y \neq X$. It would also be convenient to have $V(o) = W$, but not a single node v such that $V(o) = \{v\}$, where 'obvious' proposition is $o = "x \in \Theta"$. To progress further, one now needs to scrutinise relation R . The natural interpretation of R could be the set inclusion relationship on Θ — the support of a more general statement contributes nothing to the support of its subsets, but support of subsets increases beliefs in the more general proposition.

Definition 3.3.2 (Set W) Let Θ be a frame of discernment with support function S , let $C(S)$ be the core of S . Consider a Kripke model $\mathfrak{F}_m = \langle W, R, V, m_w \rangle$ meant to represent Θ . Language \mathcal{L} is the same as defined on page 24, where the set of variables: $\text{Var}\mathcal{L} = \{x \in A : A \subseteq \Theta\}$. Let p be a variable in language \mathcal{L} :

$$p = "x \in A"; p \in \mathcal{L}; A \subseteq \Theta.$$

Then for every set $A \in C(S)$, $A \neq \Theta$, there must be a node $w \in W$ such that $w \in V(p)$.

Relationship R is determined by set inclusion on Θ , and may be pretty easily described by the definition below.

Definition 3.3.3 (Relation R) Let $\mathfrak{F}_m = \langle W, R, V, m_w \rangle$ be a semantic model with mass assignments for frame of discernment Θ . Assume that

$$p = "x \in A"; q = "x \in B"; p, q \in \mathcal{L}; A, B \subseteq \Theta;$$

and let

$$v \in V(p), w \in V(q); \text{ where } v, w \in W$$

then

1. If $A \cap B \neq \emptyset$, then $V(p \wedge q) \neq \emptyset$;
2. If $A \cap B = \emptyset$, then $v \not R w$ and $w \not R v$, where $a R b$ means that a cannot see b ;
3. $B \subset A$ and v, w such that $w \in V(p) \setminus V(q)$ and $v \in V(q)$ requires that $v R w$ and $w \not R v$, given $A, B \notin C(S)$;

4. If there is $w \in W$ such that $w \in V(p)$, and $m_w(w) < 1$, then there must exist $v \in W$ such that $v \notin V(p)$.

The second part of condition 3 may look confusing at first glance. The requirements for the sets not to belong to the core of the support function is explained below and still stems from the need to have non-zero mass assignments for the nodes that represent core element and still to account for the inclusion relationship. In many cases condition 4 is equivalent to having $r = "x \notin A"$ and $v \in V(r)$.

Next, the probability masses are assigned to existing nodes. There is nothing too complicated about this step; the already known mass assignments of the core elements of the support function should be preserved. It is done with one exception: because of the reasons outlined above, there is no node that corresponds to the whole of the frame of discernment. Instead, a node that does not validate statements true at any other node is introduced. This node does not validate the negations of those statements either: it rather stores the unassigned mass.

Definition 3.3.4 (Node mass assignments) *Let p be as in Definition 3.3.3. Let S be a simple support function centred around A . The probability masses in the corresponding Kripke model $\mathfrak{F}_m = \langle W, R, V, m_w \rangle$, where $W = \{w_1, w_2\}$ and $V(p) = \{w_1\}$ are then defined according to the following rules:*

$$\text{if } m(A) = s, \text{ then } m_w(w_1) = s, \quad (3.8)$$

and

$$\text{if } W \setminus V(p) = w_2, \text{ then } m_w(w_2) = 1 - s. \quad (3.9)$$

The subscript w is used for m_w to stress the difference between mass assignments of elements of 2^Θ and mass assignments of the nodes in W . Even though the values coincide on the core elements of support function S , the domains of the functions are different, and they represent different things. The definition above can be extended to arbitrary support functions:

Definition 3.3.5 *Let Θ be a frame of discernment, let S be a support function over Θ and \mathfrak{F} be a semantic model representing this support function. The probability mass assignments of the nodes in \mathfrak{F} are then defined according to*

$$\forall A \in C(S); \exists w \in W, \text{ such that } m_w(w) = m(A).$$

It is important to see that in this case existence of a node whose mass is equal to the mass of element A from the core of the support function, does not mean that proposition " $x \in A$ " is valid only at this node. Rather, the node supports exactly this atomic proposition and nothing else.

To facilitate the discussion, note that a Kripke frame is a directed graph. In a directed graph the nodes that do not have any outgoing edges except for the ones pointing to itself is called a *terminal node* [87].

Only terminal nodes have non-zero mass assignments, non-terminal nodes have zero mass assignments. The support function for non-terminal nodes is the sum of mass assignments of the terminal nodes reachable from them:

$$\forall v, w \in W \text{ such that } vRw, m_w(v) = 0, \quad bel(v) = \sum_{w \in W, vRw} m_w(w). \quad (3.10)$$

Equation (3.10) gives an explanation why special amendments for the core elements in Definition 3.3.3 are needed. According to the definition above, it is enough to have a terminal node for each core element. There is no problem in case of a simple support function — there are two separate nodes, each is terminal, and each one has some mass assigned to it. In case of a more complicated support function one may run into a situation when some core elements *include* each other. In this case, the model should have a non-terminal node with its own mass assignment that may differ from the sum of the masses of terminal nodes reachable from it. Such a frame can not be a collection of disconnected nodes. In this case calculating the *beliefs* becomes impossible. There is a fairly straightforward trick that helps to overcome the hurdle without violating any of conditions stated above.

Let us first look at translating a frame with a simple support function to a Kripke model and then at a procedure of translating a frame with an arbitrary support function over it.

3.3.2 A simplest frame

As before let Θ be a frame of discernment and x be the quantity of interest. Let the value of the quantity x be within set $A \subset \Theta$, with $m(A) = s_1$, nothing else is known. The support function for this frame of discernment is

$$S(B) = \begin{cases} 0 & \text{if } B \neq A, B \neq \Theta; \\ s_1 & \text{if } B = A; \\ 1 & \text{if } B = \Theta. \end{cases}$$

According to the procedure outlined in the previous section, the minimal W is then $\{w_1, w_2\}$, relation R is empty except for the reflexive pairs. Valuation V is $V(p) = \{w_1\}$ and $W \setminus V(p) = \{w_2\}$, where p as in Definition 3.3.2. Graphical representation of the model is in Figure 3.4.



Figure 3.4: A single piece of evidence Kripke model

In the model above $m_w(w_1) = s_1$ and $m_w(w_2) = 1 - s_1$. The assignments are done under the assumption that the total should be one over the whole model: a very basic requirement for any evidential setup. These numbers are the basic probability assignments and may be used for calculating the values of the corresponding belief function.

The set of atomic propositions validated at node w_2 is left empty. This is in line with the earlier discussion about trying to avoid creating 'supernodes'. Moreover having a node that validates ' x is in Θ ' violates the conditions of Definition 3.3.3: this node cannot be a terminal node, but it must have a non-zero mass assignment. Generally, the nodes that correspond to the statement of this kind are not listed: while not giving any additional information, they must be connected to every node in a model obscuring the rest of relationships. Recall the earlier remarks about a 'supernode'. There is no contradiction to the original premise, the mass assignment of a non-terminal node is the sum of the assignments of the nodes reachable from it. The belief in the statement above should be zero and it is: as it is composed of the mass assignments $m_w(w_1)$ and $m_w(w_2)$. The belief in the statement corresponding to $m(\Theta) = 1 - s$ is the belief in any statement that is not based on the premise that A is true, so it is $m_w(w_2)$. When the mass assignments are distributed, it is important to remember that the ultimate goal of presenting frames of discernment through Kripke models is to calculate the *beliefs*, not the *mass assignments*.

3.3.3 A slightly more complicated model

One of the attractive features of the Dempster-Shafer theory is in its flexibility. In particular, the only requirement for the mass assignment is not assign any beliefs to the empty set and to have an overall sum of one. It does not require any kind of monotonicity or additivity. In most cases, especially when modelling human behaviour is concerned, such 'natural' conditions are met. On the other hand, the Dempster-

Shafer theory evidence combination rule is a more powerful procedure that handles non monotonic mass assignments as well. The example below shows a non monotonic mass assignment can be handled with the aid of the proposed interpretation.

Assume (the function is taken from [98]) the following mass assignments about possible colour of some faraway object which can only be red, white, or blue. The sensor has some confidence in determining the colours, but it also assigns some lower masses to possibility of the object being one of two or to any colour. The situation is described by the mass assignment in Table 3.4. Instead of a single centre, the support

Table 3.4: A non-monotonic mass

Hypothesis	Mass	Belief
Null	0	0
Red	0.35	0.35
White	0.25	0.25
Blue	0.15	0.15
Red or White	0.06	0.66
Red or Blue	0.05	0.55
White or Blue	0.04	0.44
Any	0.1	1.0

function is centred around several core elements. The core is $\{R, W, B, R \cup W, R \cup B, W \cup B, R \cup W \cup B\}$, every core element is represented by an atomic proposition like $x \in R \cup W$ etc. Using the already familiar notation the support function S is given below:

$$S(X) = \begin{cases} 0 & \text{if } X \notin C(S); \\ m(R) = 0.35 & \text{if } X = R; \\ m(W) = 0.25 & \text{if } X = W; \\ m(B) = 0.15 & \text{if } X = B; \\ m(R) + m(W) + m(R \cup W) = 0.66 & \text{if } R \cup W; \\ m(R) + m(B) + m(R \cup B) = 0.55 & \text{if } R \cup B; \\ m(W) + m(B) + m(W \cup B) = 0.44 & \text{if } W \cup B; \\ 1 & \text{if } X = \Theta. \end{cases}$$

Every core element must have a node assigned to it. However, this S cannot be represented by seven disconnected nodes, because some of the core elements *include* some other core elements and according to Definition 3.3.3 must see each other (if they were not core elements). Moreover, if this condition is met, then some non-terminal nodes must have a non-zero mass which also contradicts the rules introduced earlier.

A simple trick helps to avoid both of the problems above. Ignore the masses at first, assign a separate node for every core element. On the picture every node is labelled with the core elements rather than with corresponding propositional variables. To save space in the picture, the core element $R \cup W \cup B$ is labelled Θ . Copy the frame using the map $\varepsilon(x) = y, xRy, yRx \forall x \in \Theta$, reflect the result of this operation on the Kripke model. Finally, induce R according to set inclusion, but only between nodes in the first and the second rows. Formally, it means that new edges on the model will only be added between non-terminal nodes (i.e. the 'ends' of the edges).

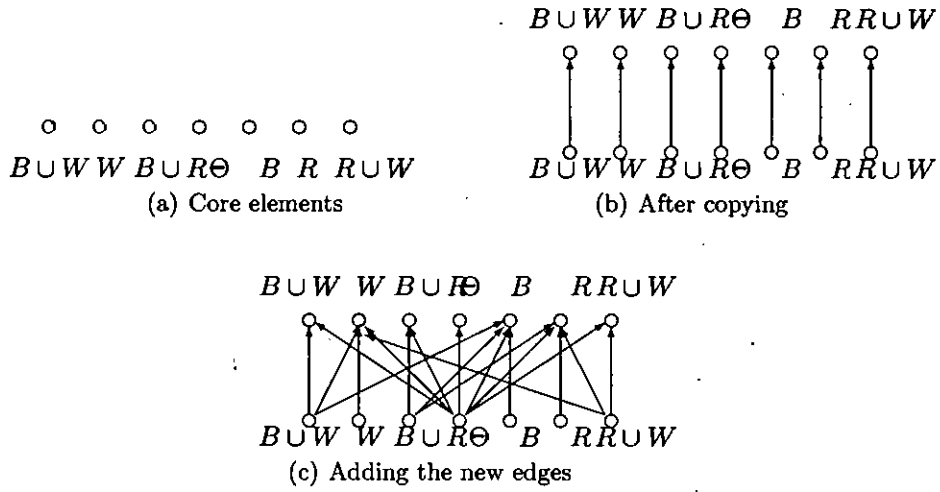


Figure 3.5: Building a model for a non-monotonic mass assignment

The steps outlined in previous paragraph are in Figure 3.5. At first, a node is assigned to each core element. This assignment produces a model in Figure 3.5(a). The only difference between this model and the models that were considered earlier (recall disjointed nodes in examples about evidence combination) is that all the elements represented by the nodes belong to the same frame of discernment. On the other hand, the model does not yet give full information that can be retrieved from the frame of discernment — no subset/ superset relationships are shown, and thus the only conclusions that can be drawn from the model at the moment are the degrees of confidence attributed to the core elements.

Picture 3.5(b), shows what happens if the nodes of the original frame of discernment are copied. It is important to note that even though in terms of sets nothing has changed, the semantic of the model *has* changed. Kripke model on Figure 3.5(a) is not equivalent to the one on Figure 3.5(b). A model that contains only single nodes verifies all the formulac in Boolean logic Cl. Recall that the model in Figure 3.5(b)

refutes the law of excluded middle. The goal is to build a logic that allows 'strict' inference without resorting to non-constructive proofs, so the process is on the right track so far.

The last step is shown in Figure 3.5(c). The only nodes that gained new edges were the nodes that were generated at the very first step. In Kripke model's terms, this condition means that only non-terminal nodes can gain new edges, and that there are no edges between terminal nodes. This condition might seem a little counter intuitive, but it only appears here because of the nature of the mass assignment since the mass assignments of the core elements describe the beliefs attributed *exactly* to them. It is also known that the beliefs that are put upon non-terminal nodes are made of the masses of assigned to the terminal nodes that can be reached from them. In this case, the mass assigned to the event itself (say object being Red or White) is smaller than the masses of its subsets. The small confidence in the event itself is reflected in the low mass assigned to corresponding terminal node. Higher belief and plausibility of the same event are reflected in the sums of the masses of the nodes that can be reached from the corresponding non-terminal node.

The terminal node corresponding to the same event occurs *later* than the non-terminal one. Semantically, it means that even though one can probably verify more formulae at the terminal node, it can only be done with confidence such that belief=plausibility=mass. There are fewer formulae validated at the earlier node, but there is higher confidence put on them: belief is no longer equal to plausibility.

3.3.4 Updating the evidence

The two examples demonstrate how a simple support function and an arbitrary (non-separable, non-monotone) support functions can be translated to Kripke models. This demonstration gives us enough insight to proceed towards a more interesting task—updating the evidence.

There are two possible situations — homogeneous and heterogeneous evidence. Heterogeneous evidence involves two different support functions that put some confidence on different statements. If supported statements overlap, then there is non-contradicting confidence, if the statements do not overlap, that means that different pieces of evidence contradict each other. Homogeneous evidence means that there are different support functions that support exactly the same statements.

Heterogeneous evidence may be thought of as the situation when the same decision maker learns new facts. The homogeneous case corresponds to the situation when

two or more different decision makers describe their beliefs attributed to the same facts.

The latter situation gives less room for inference. The only update that may happen will change the degrees of confidence attributed to the statements known to be true. In other words, it does not change the set of true formulae, but updates their support.

In case of heterogeneous evidence combination, the set of true formulae changes. From the semantic point of view the heterogeneous case is more promising. Even though the resulting semantics is more complicated, the procedure for heterogeneous evidence update is simpler, or, better said, less awkward. Heterogeneous evidence combination is illustrated on Shafer's famous examples already considered in the introduction to the Dempster-Shafer theory part of the chapter.

Homogeneous evidence combination

Consider two elementary mass assignments (simple support functions)

$$m_1(A) = s_1; \quad m_1(\Theta) = 1 - s_1,$$

$$m_2(A) = s_2; \quad m_2(\Theta) = 1 - s_2.$$

Both mass assignments support exactly the same proposition, but give rise to two different support functions that reflect different degrees of confidence in statement attributed to set A . Both functions are of the same form given below:

$$S_i(B) = \begin{cases} 0 & \text{if } B \neq A; B \neq \Theta, \\ s_i & \text{if } B = A, \\ 1 & \text{if } B = \Theta, \end{cases}$$

where $i = 1, 2$. This situation can be interpreted as two decision makers expressing their opinion on the same matter. Updating the evidence results in a certain common confidence level. There is nothing new to this situation — both pieces of evidence are combined the same way as was already demonstrated in relevant sections.

The intersection table for these two mass assignments is in Table 3.3.4. Since both mass assignments support the same propositions, the new mass assigned to A is the same as the mass assigned to its subsets. The table gives:

$$m_1 \oplus m_2(A) = s_1 s_2 + s_1(1 - s_2) + s_2(1 - s_1) = s_1 + s_2 - s_1 s_2,$$

Table 3.5: Homogeneous evidence intersection table

$m_1 \oplus m_2$	$A \ s_1$	$\Theta \ (1 - s_1)$
$A \ s_2$	$A \ s_1 s_2$	$A \ s_2(1 - s_1)$
$\Theta \ (1 - s_2)$	$A \ s_1(1 - s_2)$	$\Theta \ (1 - s_1)(1 - s_2)$

and

$$m_1 \oplus m_2(\Theta) = (1 - s_1)(1 - s_2).$$

It was already shown that such mass assignments give a simple support function of the form

$$S_{m_1 \oplus m_2}(B) = \begin{cases} 0 & \text{if } B \neq A; B \neq \Theta, \\ s_1 + s_2 - s_1 s_2 & \text{if } B = A, \\ 1 & \text{if } B = \Theta. \end{cases}$$

A simple support is represented by a simplest two-node model. The only point that is not entirely clear is how this two-node model is produced as a result of combining two original models that are constructed unambiguously. A little analysis shows that there is nothing especially tricky about the situation. The only drawback of such evidence combination is its apparent awkwardness that requires a few steps to arrive exactly to the point where it started bar the mass assignments. Original semantics is represented by two two-node models shown in Figure 3.6(a), since the evidence is supported by two different mass assignments the masses assigned to the nodes are listed next to them. The intersection table 3.3.4 induces the model given in Figure 3.6(b), the new nodes are labelled with the sets they represent (their mass assignments are not known yet). After analysing the model in the picture one can see that some nodes are actually equivalent (they have the same incoming and outgoing edges), so they can be collapsed into single nodes resulting in model given in Figure 3.6(c), the nodes now are labelled with the sets membership in which they validate. The last model is different from the model for a single support function earlier with the new mass assignments listed next to the nodes. The model in Figure 3.6(c) has a 'supernode', but this situation was already discussed and semantically the model is equivalent to the one in Figure 3.6(d).

3.3.5 Building the models

Definition 3.3.3 is used to construct the Kripke models that represent the frames of discernment describing the examples from Section 3.1.2.

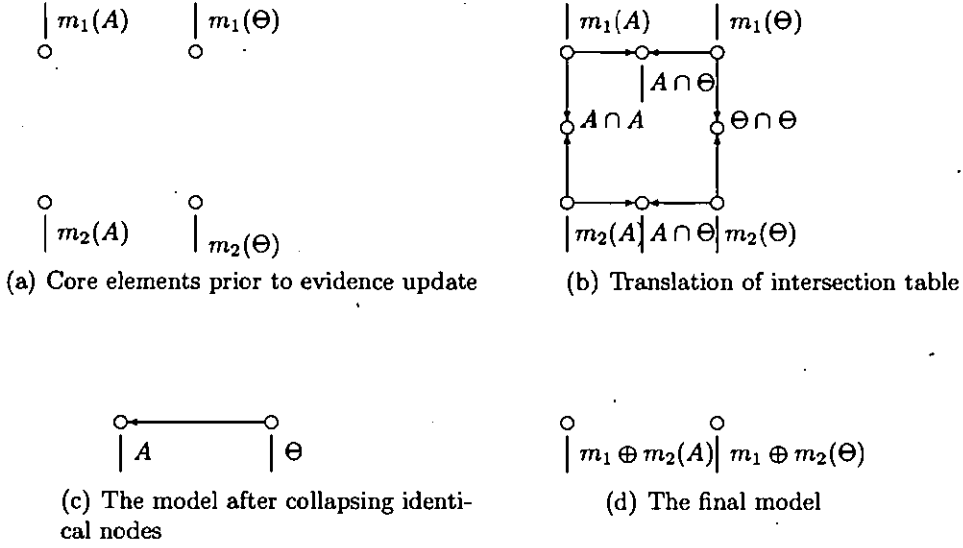


Figure 3.6: this is homogeneous evidence

Burglary of sweetshop

The story behind this example is found on page 36. Recall, that known facts may be represented by two elementary frames $\Theta_1 = \{L, R\}$ and $\Theta_2 = \{I, O\}$ that are later combined into a universal frame of discernment $\Theta = \{LI, LO, RI, RO\}$ and two different probability assignments over its subsets.

First piece of evidence tells the thief is a left-handed person $m_1(\{LI, LO\}) = s_1, m_1(\Theta) = 1 - s_1, m_1(\text{elsewhere}) = 0$. The second piece of evidence tells that the thief is an insider $m_2(\{LI, RI\}) = s_2, m_2(\Theta) = 1 - s_2, m_2(\text{elsewhere}) = 0$. Even though there is only one frame of discernment and two pieces of evidence over it, more formal way to look at the same situation includes *two* separate frames of discernment with *two* separate mass assignments combined into Θ after evidence update.

The same distinction works for Kripke models: prior to the evidence combination there are two elementary models, similar to the ones in Figure 3.4, that are later combined into a new, more complicated one. The initial setup is represented in Figure 3.7. The first frame is $\Theta_1 = \{L, R\}$ following the notation abuse introduced earlier, induces nodes w_1, w_2 . The second frame is $\Theta_2 = \{I, O\}$ gives nodes w_3, w_4 . The relation R is empty except for reflexive pairs. It is the situation when two separate

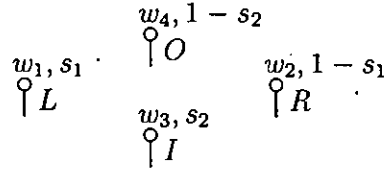


Figure 3.7: Burglary of sweetshop, initial setup

pieces of evidence are presented but not combined. To update the evidence one has

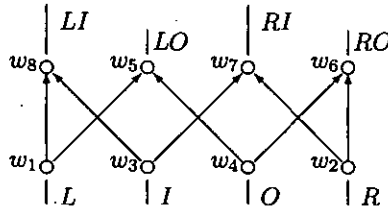


Figure 3.8: Final frame for burglary of sweetshop

to combine two separate models. This is done by looking at the possible intersections of elements of Θ_1 and Θ_2 and then constructing R according to rules (1)–(4) from Definition 3.3.3. The resulting model shown in Figure 3.8. To preserve the readability, the probability assignments for the points are not listed on the picture, instead, they are in Table 3.6.

Before looking at the probability mass assignments in the new model, the following should be considered. When the data is combined, a new probability assignment function is constructed. The new function has the core different from the core of the original mass assignment functions. The terminal nodes in the new model are different from the terminal nodes of the original models. To get the mass assignments for this new function assume that given two pieces of evidence with basic probability assignments $m_1(A) = s_1$ and $m_2(B) = s_2$, the updated belief in $A \wedge B$, where $A \wedge B$ is an obvious shorthand for ' $x \in A$ ' AND ' $x \in B$ ' (this abbreviation will be used in the following sections), is given by $m(A \wedge B) = s_1 s_2$ and that the proposition $A \wedge B$ merits a new, terminal node. Shafer used the same postulate for developing the evidence combination rule. Keeping the above in mind and using the steps outlined in Definitions 3.3.3 and 3.3.4, the procedure yields the results given in Figure 3.8 and Table 3.6.

When the graphical representation of the model is complete, the sum of basic probability assignments is checked. The basic probabilities assigned to the terminal nodes add up to one. The degree of support for any of non-terminal nodes is retrieved by adding the probability assignments of the terminal nodes that could be reached from them.

Table 3.6: Probability assignments for Burglary of Sweetshop

Node w_i	$m_w(w_i)$	Node w_i	$m_w(w_i)$
Terminal Nodes			
w_5	$s_1(1 - s_2)$	w_7	$(1 - s_1)s_2$
w_6	$(1 - s_1)(1 - s_2)$	w_8	s_1s_2
Non-terminal Nodes			
w_1	$m_w(w_5) + m_w(w_8) = s_1$	w_3	$m_w(w_7) + m_w(w_8) = s_2$
w_2	$m_w(w_6) + m_w(w_7) = 1 - s_1$	w_4	$m_w(w_5) + m_w(w_6) = 1 - s_2$

In Table 3.6 we introduce yet another slight abuse of notation: only the terminal nodes have probability masses assigned to them; the non-terminal nodes have some beliefs equal to the sum of masses of the terminal nodes reachable from them. However, these are not exactly the same as the beliefs discussed in the Dempster-Shafer portion of the work. For the lack of a better term and to keep the notation consistent these values are still called masses and denoted $m_w(w_i)$. For example, the belief that a thief is left-handed is now equivalent to belief in $m_w(w_1) = m_w(w_5) + m_w(w_8) = m((LI) \vee (LR)) = s_1s_2 + s_1(1 - s_2) = s_1$. It is the initial belief attributed to a left-handed suspect. The created model does not change the initial beliefs. It rather gives beliefs attributed to some new statements. The next example shows when and how initial beliefs may be updated.

Cabbage seed

The available knowledge can be represented by two frames of discernment $\Theta_1 = \{A, \neg A\}$ and $\Theta_2 = \{B, \neg B\}$. The mass assignments for the elements of these frames of discernment were given when the example was first analysed on page 37. Assigning nodes is also straightforward: both frames of discernment are represented by two nodes that do not see each other. Having two isolated nodes is the case whenever only one statement is supported. Such models show the situation when a decision maker has only two options — either to believe in the statement or in its negation.

The Kripke model representing the example prior to the evidence combination is identical to the one in the burglary of the sweetshop. The model after the evidence combination has fewer nodes than the sweetshop's: the intersection Table 3.2 shows that since one of the pieces includes another there are only three terminal nodes. Figure 3.9 shows the model corresponding to the cabbage seed example.

In Figure 3.9, note that node w_1 is not a terminal node anymore, even though

it was in the original model. The opposite is true for node w_3 : it stayed a terminal node even after the evidence was combined. Nodes w_5 and w_6 are the ones that were created *after* the evidence was combined. What happens to nodes w_1 and w_3 is explained by looking at the intersection table. In the formulae below, V_1 and V_2 are the valuation functions of the models prior to the evidence combination, and V is the valuation function for the model given in Figure 3.9:

$$V_1(A) = \{w_3\}, \quad V_1(\neg A) = \{w_4\},$$

$$V_2(B) = \{w_1\}, \quad V_2(\neg B) = \{w_2\}.$$

Checking the intersections of the pieces of evidence gives us the following set of statements that merit a terminal node: $A \wedge B = A$, $A \wedge \neg B = \emptyset$, $\neg A \wedge B$, and $\neg A \wedge \neg B$. The model has only three terminal nodes, one of which, w_3 , is inherited from the original model. The corresponding beliefs are in Table 3.7. A straightforward

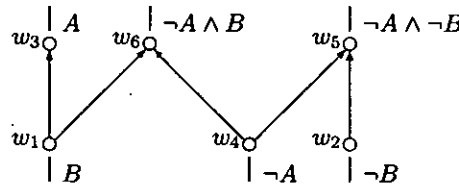


Figure 3.9: Cabbage seed

check shows that the beliefs attributed to the terminal nodes add up to one and thus no normalisation is needed. Calculating the corresponding belief function gives the same beliefs as in Shafer's example. The model is a truthful representation of the classical example.

Table 3.7: Probability assignments for Cabbage Seed

Node w_i	$m_w(w_i)$	Node w_i	$m_w(w_i)$
Terminal Nodes			
w_3	c_1	w_5	$(1 - c_1)(1 - c_2)$
w_6	$(1 - c_1)c_2$		
Non-terminal nodes			
w_1	$m_w(w_3) + m_w(w_6) = c_1 + c_2 - c_1c_2$	w_2	$m_w(w_5) = (1 - c_1)(1 - c_2)$
w_4	$m_w(w_5) + m_w(w_6) = 1 - c_1$		

To see how combining two frames of discernment affects beliefs, consider the total

belief attributed to node w_1 . From Table 3.7 the sum of mass assignments of the nodes reachable from w_1 is

$$m_w(w_1) = m_w(w_3) + m_w(w_6) = c_1 + (1 - c_1)c_2 = c_1 + c_2 - c_1c_2,$$

which, along with $c_2 > c_1$, tells us that the resulting belief in proposition B has gone up. At the same time, the support for proposition A stays unchanged.

Alibi

Start again with two elementary 2-node frames: $\Theta_1 = \{G, \neg G\}$ and $\Theta_2 = \{I, \neg I\}$. The resulting frame after the evidence combination has nodes for all possible intersections between elements of Θ_1 and Θ_2 . There are more such nodes than in the cabbage seed case. Two pieces of evidence contradict each other, so no piece of evidence from one frame can include another one. The statements like '*not guilty AND innocent*' make perfect sense and thus result in new nodes. The only impossible intersection is then '*innocent AND guilty*', which is missing from the resulting model. The model after the evidence combination is shown in Figure 3.10. The model is different from

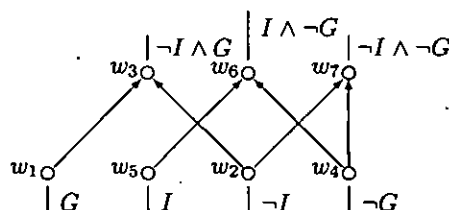


Figure 3.10: Alibi frame

the two preceding ones: the mass assignments for the terminal nodes do not add up to one anymore, so some normalisation procedure is needed. The normalisation follows Shafer: the probability assignments that do not support anything, the beliefs committed to \emptyset are ignored. The basic probability attributed to the empty set is given by s_1s_2 , so the normalising factor is $1 - s_1s_2$. The normalised basic probability assignments are in Table 3.8.

There is also a non-obvious node w_7 . This node is needed because having two non-overlapping pieces of evidence does not necessarily mean that their complements do not intersect. Intuitively, the situation at w_7 can be explained as a kind of state of ignorance, believing neither piece of evidence. There is no node corresponding to the empty set though. The total of available probability assignments now amounts to one, and that the corresponding support function can be retrieved easily.

Table 3.8: Alibi Probability Assignments

Node w_i	$m_w(w_i)$	Node w_i	$m_w(w_i)$
Terminal nodes			
w_3	$\frac{s_1(1-s_2)}{1-s_1s_2}$	w_6	$\frac{s_2(1-s_1)}{1-s_1s_2}$
w_7	$\frac{(1-s_1)(1-s_2)}{1-s_1s_2}$		
Non-terminal nodes			
w_1	$m_w(w_3) = \frac{s_1(1-s_2)}{1-s_1s_2}$	w_2	$m_w(w_3) + m_w(w_7) = \frac{1-s_2}{1-s_1s_2}$
w_4	$m_w(w_6) + m_w(w_7) = \frac{1-s_1}{1-s_1s_2}$	w_5	$m_w(w_6) = \frac{s_2(1-s_1)}{1-s_1s_2}$

3.4 Verification of the approach

So far, it has been demonstrated that the procedure outlined in Section 3.3 helps to translate Shafer's popular examples to the language of Kripke models, and that combining two models according to rules of Definition 3.3.3 induces the same beliefs over the resulting model as produced according to the Dempster-Shafer evidence combination rule. In this section, it is shown that the approach always works, and that it provides a meaningful translation of the frames of discernment to semantic models. The proof is done through showing that the basic probabilities assigned to the terminal nodes of the constructed models always give rise to a belief function, and that the rules for combining two models provide an analogue to Dempster-Shafer evidence combination rule.

The observations above should be stated as a proposition.

Proposition 3.4.1 *Given a support function S on a frame of discernment Θ , a Kripke model with mass assignments on nodes \mathfrak{F} satisfying the conditions of Definitions 3.3.2, 3.3.3 and 3.3.4 unambiguously represents support function S in a sense that given a set $A \subseteq \Theta$ and a proposition*

$$p = 'x \text{ is in } A'$$

where x is some quantity of interest we have

$$S(A) = \sum_{w \in V(p)} m_w(w).$$

Conversely, given two models \mathfrak{F}_1 and \mathfrak{F}_2 representing two support functions S_1 and S_2 they can be combined into a new model \mathfrak{F} satisfying conditions of Definitions 3.3.2, 3.3.3 and 3.3.4 and representing support function S defined through mass assignment $m = m_1 \oplus m_2$.

Proof: The part about the belief functions is straightforward. Set W and relation R were constructed according to rules (1)-(4) from Definition 3.3.3. By construction, $m_w(w_i)$ satisfies conditions (1)-(2) of Definition 2.1.1, and thus gives rise to a belief function. The constructed models thus represent Dempster-Shafer evidential frameworks.

Recall that $\forall A \subseteq \Theta$ the belief is given by $Bel(A) = \sum_{B \subseteq A} m(B)$. In a model $\mathfrak{F} = \langle W, R, V, m_w \rangle$ for any node $v \in W$, the belief in statements valid at this node is $Bel(v) = \sum_{w \in W, vRw} m_w(w)$. The nodes are ordered according to set inclusion on Θ . Every terminal node corresponds to an element of $\mathcal{C}(S)$, every non-terminal node can see some terminal nodes. Thus, for any set $A \subseteq \Theta$ and for the statement “ $p = x$ is in A ”, we have that $V(p)$ includes all the nodes representing core elements that are inside of A with masses assigned according to Definition 3.3.4. But this means that the beliefs induced over the model are the same as calculated directly from the frame of discernment.

The part about the evidence combination rule is a little more involved. The proof will be done by considering the properties of set W and relation R in the resulting model.

First, consider the initial separate models. As there are two different models, there is a temptation to consider two separate frames of discernment. It is not the case: if there are two separate frames of discernment, then the evidence represented by one is not always combinable with the evidence from the other. Instead, assume that the frame of discernment is large enough to accommodate all the claims and distinguish between sets W_1 and W_2 . There are no requirements to sets W_1 and W_2 yet. The procedure described above dealt with the simplest type of evidence: the one that supports some statements from \mathcal{L} and the complement of their union. In the analysed examples the beliefs were mostly binary, as in the example of the left- or right-handed thief. Imagining three-handed creatures performing theft in the same example leads to assigning beliefs to three statements. The binary requirement is not an absolute must.

In the simplest case, the models consist of disconnected nodes. This structure makes relation R empty except for the reflexive pairs. Consider two models $\mathfrak{M}_1 = \langle W_1, R_1, V_1 \rangle$ and $\mathfrak{M}_2 = \langle W_2, R_2, V_2 \rangle$ based on the pieces of evidence that are combined. Let $A, B \in \Theta$ be some subsets of the frame of discernment, θ be the quantity of interest; $p = ‘\theta \in A’$ and $q = ‘\theta \in B’$ be the statements of \mathcal{L} , and $m_i(X)$, $X \in \Theta$, $i = 1, 2$ be the known basic probability assignments.

The elements of respective sets W_i are determined according to the available ev-

idence for each piece. If $m_i(A) > 0$, then $\exists w_i^j \in W_i$ such that $V_i(p) = w_i^j$ and $m_w^i(w_i^j) = m_i(A)$.

On W_i , the condition $\sum_{w_i^j \in W_i} m_w^i(w_i^j) = 1$ is satisfied by construction according to rule (4) from Definition 3.3.3, and the basic probability assignment for a statement's complement is given in equation (3.9). Moreover, translating the initial evidential setup produced the models in which every node is a terminal node.

When models M_1 and M_2 are combined, a new model $M = \langle W, R, V, m_w \rangle$ is generated. Set W of the new model contains all the nodes from W_1 and W_2 , plus it has new nodes reflecting the combined evidence.

Assume that propositions p and q are as above, and that $w_1^A \in V_1(p)$ and $w_2^B \in V_2(q)$; assume also that $A \cap B \neq \emptyset$. The new model then has nodes $w^A \in V(p)$, $w^B \in V(q)$, and a new node, $w^{A \cap B}$ such that $\{w^{A \cap B}\} = V(p \wedge q)$. Semantically, $w^{A \cap B}$ is the node where the proposition ' $\theta \in A \cap B$ ' is true. According to Definition 3.3.3, the new node is in the following relation to its 'ancestor' nodes $w^A R w^{A \cap B}$, $w^B R w^{A \cap B}$. It is a terminal node whose basic probability assignment is $m_w(w^{A \cap B}) = m_w(w^A)m_w(w^B)$.

The situation above is the most straightforward case illustrated on the burglary of the sweetshop example. Definition 3.3.3 describes the other cases as well. The cases when $A \cap B = \emptyset$ or $A \subset B$ result in keeping some of the nodes from \mathfrak{M}_i in terminal positions and keeping their original basic probability assignments. Such 'preserved' assignments need a little purely formal clarification.

Let $v \in W_1$ be such a node, let $m_w^1(v)$ be its basic probability assignment in W_1 . In updated model \mathfrak{M} this node has the same probability assignment $m_w(v) = m_w^1(v)$. The 'preserved assignment' can be seen as $m_w(v) = m_w^1(v) \cdot 1 = m_w^1(v)m_2(\theta \in \Theta)$ to stress the parallelism with the Dempster-Shafer rule for subsets, which always operates with pairs of probability assignments. One also needs to remember that probability assignments of all non-terminal nodes are zeros.

It is now time to check whether the conditions of Definition 3.1.2 are satisfied. Equation (3.3) is always satisfied. Verification is a matter of translating between the languages of subsets and Kripke models. The probability assignments of the subsets of Θ are now represented by the probability assignments of the nodes of W . Both models \mathfrak{M}_1 and \mathfrak{M}_2 satisfy $\sum_{w_i^j \in W_i} m_w^i(w_i^j) = 1$. Given the correspondence between subsets of Θ and elements of W_i , define the set

$$C = \{w, v \in W \mid w R v\}.$$

$C \subseteq W$, with $C = W$ only when the two pieces of evidence flatly contradict each

other. The case of evidence that is impossible to combine was considered and ruled out by Shafer. In all other cases, there is a proper set inclusion. Whenever inclusion $C \subseteq W$ is proper, the inequality

$$\sum_{\substack{wRv; \\ w \in W_1, v \in W_2}} m_w^1(w)m_w^2(v) < \sum_{w \in W_1} m_w^1(w) \sum_{v \in W_2} m_w^2(v) = 1$$

is true, so equation (3.3) is satisfied.

Equation (3.4) is the normalising factor. The normalising factor for model \mathfrak{M} is calculated according to

$$N^M = 1 - E = 1 - \sum_{wRv} m_w^1(w)m_w^2(v) = \sum_{wRv} m_w^1(w)m_w^2(v),$$

where the summation is taken over all the possible pairs $w \in W_1$ and $v \in W_2$. Recalling the earlier remark about the nodes that stay terminal in the combined model, one can see that the normalising factor used in the interpretation is the same as in equation (3.4).

Checking the conditions of equations (3.5) and (3.6) is now straightforward. Equation (3.5) is trivially satisfied: according to Definition 3.3.3, no node is in $V(\emptyset)$, and thus no basic belief is attributed to the empty set. In model \mathfrak{M} the probability assignment of proposition p is calculated by

$$m(p) = \frac{1}{N} \sum_{w^i \in V(p)} m_w(w^i),$$

where w^i 's are the nodes within the valuation of p . The probability assignments of any terminal node $w^k \in W$ are based on the probability assignments of its ancestors $w_1^i \in W_1$, $w_2^j \in W_2$ using

$$m_w(w^k) = m_w^1(w_1^i)m_w^2(w_2^j)$$

and that $m_w(v) = 0$ whenever $v \in W$ is not a terminal node. This effectively means that whenever $V(A) = \{w\}$ we have that $m(A) = m_w(w)$, where the left-hand side refers to the probability assignments over frame Θ and the right-hand side to the probability assignments over set W . The parallelism works whenever the evidence represented by different probability assignments is combinable. Thus the probability assignment m over W is indeed the orthogonal sum of basic probability assignments m_w^1 and m_w^2 over W_1 and W_2 as desired. \square

Thus, any support function can be represented by a Kripke model. It would also

be useful to see whether the converse is true: does it mean that any Kripke model with mass assignment on nodes represents as a support function. The answer is yes in case if the masses assigned to terminal nodes of a model represent core elements of some support function. The only requirement to a Kripke model to represent a support function is to have some terminal nodes. This requirement only amounts to excluding the cyclical constructions, that are semantically equivalent to single nodes (in case of transitive nodes).

Proposition 3.4.2 *Let $\mathfrak{K} = \langle W, R, V, m_w \rangle$ be a Kripke model with mass assignments on nodes. Then there is a support function S over frame Θ , such that for any $X \in \mathcal{C}(S)$ there is a node $w \in W$ such that $m_w(w) = m(X)$, and that for any set $A \subset \Theta$ we have that $S(A) = \sum_{w \in V(\{x \in A\})} m_w(w)$, where x is some quantity of interest.*

Proof: The first half is obvious: every terminal node corresponding to a statement $x \in X$ is some element in $\mathcal{C}(S)$. The second part is also automatic: relation R is a partial order; the ordering of subsets by set inclusion on Θ is also a partial order. Therefore, it is possible to find a subset A of Θ such that $S(A) = \sum_{w \in V(\{x \in A\})} m_w(w)$. \square

The results above demonstrate that any Kripke model can correspond to a Dempster-Shafer support function and vice-versa. This means that a formula valid in some intuitionistic Kripke model corresponds to some situation described by the means of the Dempster-Shafer theory. This observation also points to the minimal intuitionistic logic **Int**. To see why it is true, it is useful to abstract from the definition of a logic through a calculus and look at a logic as a set of true formulae. From this angle, Boolean logic **Cl** is the set of formulae valid in a single node intuitionistic Kripke model, and the minimal intuitionistic logic **Int** is the set of all formulae in language \mathcal{L} valid in all intuitionistic frames [48]. As it was demonstrated that *any* intuitionistic Kripke model with mass assignments corresponds to a support function, the set of formulae validated in our case is the latter. The Dempster-Shafer theory is represented by the minimal intuitionistic logic. On one side, the result may be viewed as disappointing: there are many stronger logics. On the other hand, the logic in question is complete and sound and certainly provides a usable reasoning apparatus.

3.5 Embedding issues

So far we have not any explicit restrictions on the semantic models that represent support functions. It does not mean, however, that we cannot benefit from numerous

embedding results available for intermediate logics [48]. Similar but non-universal results are possible for models that represent different support functions. The non-universal nature of the results should not deter one from exploring the question. Placing a particular model within a class of models that represent some well-established intermediate logic can make inference over that model more effective. Knowing inclusion relationship with respect to systems admitting finite axiomatisation allows one to use their syntax for inference within the analysed setup.

A few simple embedding observations should give the taste of the problem. $\text{Int} \subseteq L \subseteq \text{Cl}$, where L is any intermediate logic [48]. This interval is quite broad: there are infinitely many intermediate logics. To get non-trivial results one can look into particular si-logics. Any connected frame representing a support function in refutes Dummett's formula $(p \rightarrow q) \vee (q \rightarrow p)$, and in this case the inference can exclude tautology $(p \rightarrow q) \vee (q \rightarrow p)$ at some nodes. There are a few more general axioms that are satisfied depending on the number of evidence combinations performed and on possible branchings within each piece of evidence.

The following definitions are needed for the later discussion.

Definition 3.5.1 (Chains) *A chain of length n in a frame $\mathfrak{F} = \langle W, R \rangle$ is a set $C \subseteq W$ such that for any $v, w \in C$ either vRw or wRv must be true. The length of a chain is the number of elements in it. The depth of a frame is determined by its longest chain. An antichain is a set $A \subseteq W$ such that for any $v, w \in A$ both $v \not R w$ and $w \not R v$ are true. The width of a frame is determined by its longest antichain. The branching of a frame is the maximal number of distinct immediate successors of a node.*

The family of axioms bw_n is a generalisation of Dummett's formula (with renamed variables): $\text{bw}_n = \bigvee_{i=0}^n (p_i \rightarrow \bigvee_{j \neq i} p_j)$, $n \geq 1$. The corresponding logic is BW_n . bw_n is validated by a frame if its every rooted subframe is of width $\leq n$. The width of rooted subframes in Kripke models induced by evidence combination depends on the number of subsets within every piece of evidence and by the number of times the evidence combination rule was applied. The majority of cases described by the Dempster-Shafer theory deals with finite universes, making it possible to select n large enough to satisfy bw_n .

A similar argument works for the family of axioms bd_n , where $\text{bd}_1 = p_1 \vee \neg p_1$ is a familiar law of excluded middle or (A10) in a slightly modified notation. A member of this family of axioms is given by the recursive formula $\text{bd}_{n+1} = p_{n+1} \vee (p_{n+1} \rightarrow \text{bd}_n)$. The corresponding family of logics is BD_n . A frame is known to refute bd_n if it has

a chain of length $n + 1$. The length of a chain in a frame within \mathcal{C}_{DS} depends only on the number of times the evidence combination was applied.

The family of axioms $bb_n = \bigwedge_{i=0}^n ((p_i \rightarrow \bigvee_{j \neq i} p_j) \rightarrow \bigvee_{j \neq i} p_j) \rightarrow \bigvee_{i=0}^n p_i$, $n \geq 1$ is validated by finite frames of branching $\leq n$ giving rise to a family of logics T_n . The branching of a frame in \mathcal{C}_{DS} is determined by the maximal number of subsets within each piece of evidence. Thus there is a fairly straightforward criterion for this family of axioms too.

Looking at the conditions above, one can construct the conditions for a Kripke frame \mathfrak{F} satisfying all three axioms for an appropriate $n \in \mathbb{N}$. This can be written as $\text{Int} + bw_n \wedge bd_n \wedge bb_n \subseteq \text{For}\mathfrak{F}$, where $\text{For}\mathfrak{F}$ is a set of formulae valid in \mathfrak{F} . The same line of reasoning suggests that if the estimate for n was accurate then for $n - 1$ the relationship is $\text{For}\mathfrak{F} \subseteq \text{Int} + bw_{n-1} \wedge bd_{n-1} \wedge bb_{n-1}$. This approach allows one to place $\text{For}\mathfrak{F}$ in the interval $[\text{BW}_n \cup \text{BD}_n \cup \text{T}_n, \text{BW}_{n-1} \cup \text{BD}_{n-1} \cup \text{T}_{n-1}]$.

Proceeding in the same fashion, the set of formulae available for reasoning in each particular case can be significantly transformed thus making the inferential apparatus at disposal of the decision maker even more effective.

3.6 Incorporating the ‘empty’ evidence

The construction above is based on the premise that $p \vee (p \rightarrow \perp)$ is refuted in most of semantic models. This section provides a simple example that illustrates how attributing unconditional belief to the formula $p \vee (p \rightarrow \perp)$ can violate even the basic intuition.

Recall the burglary of the sweetshop example from page 36. The corresponding Kripke model was constructed on page 57. The constructive premises are violated by allowing Sherlock Holmes to incorporate empty evidence along with the one for which he has support. So, instead of having some belief in hypothesis about a left-handed thief, let us assume that no evidence whatsoever about the burglar being left-handed is available, $s_1 = 0$. Is it reasonable in this case to just assume that a person was definitely a right-handed person? From the constructivist point of view the answer is ‘no’. $p \vee (p \rightarrow \perp)$ is true only if p or $\neg p$ is decidable. For any other logic that accepts A10 the answer is ‘yes’.

Let us now leave the constructivist shell and try to incorporate the empty evidence into the model. The question then becomes: should one assume that whenever $V(p) = \emptyset$ and whenever w is a terminal node such that $w \in V(\neg p)$ that $m_w(w) = 1$? Accepting the assumption corresponds to the model in Figure 3.11. The basic proba-

bility assignments for terminal nodes RI and RO are the same as the assignments for the nodes I and O . In other words, adding an extra node R does not change anything. Such a situation does not contradict our basic intuition; introducing ‘empty evidence’ does not change anything. Moreover, the model in Figure 3.11 is the only possible

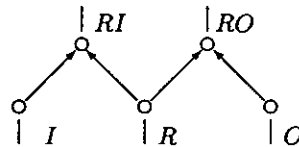


Figure 3.11: Burglary of sweetshop with empty evidence

result of combining a piece of non-conclusive evidence with whatever other evidence is available. We illustrate this observation with another Shafer’s example.

Consider now the alibi example, and assume that the second witness does not provide any information in support of the suspect’s innocence. If $p \vee (p \rightarrow \perp)$ is true, then the corresponding model is identical to the one in Figure 3.11. Two situations, quite different in non-degenerate case, become the same if one of the pieces of evidence is reduced to ignorance. Such a situation does not lead to a paradox or to an incorrect probability combination. Yet it is better avoided as it enables derivation that is ultimately vacuous and increases the complexity of the model without increasing the amount of knowledge it represents.

The ultimate goal of the current work is to develop an effective reasoning apparatus that can be implemented. From this point of view, a possibility to increase the complexity of a model without gaining anything in terms of inferential power is to be avoided.

Chapter 4

Updating knowledge in the Dempster-Shafer theory

The procedure linking the Dempster-Shafer theory frames of discernment with the semantic models was established. It was also shown that the procedure preserved the structure of the underlying frame of discernment and gave rise to the same belief function as Shafer's. In this chapter, the properties of the proposed framework are explored and further parallels between semantic models and frames of discernment are drawn.

The Dempster-Shafer theory provides a set of tools for incorporating possible changes in knowledge. These operations are different from inference on already built models. The aim of this chapter is to understand the effects of changing the frames of discernment on the corresponding Kripke models. To make the exposition more consistent, the Dempster-Shafer theory concepts and examples are given first, and then the connection with Kripke models is exposed. For brevity, we write in this chapter $V(A) \ni w$ that should be understood as $V(p) \ni w$, where $p = 'x \in A'$. We continue using notation $V(\phi)$ for truth sets of formulae.

4.1 Refining

There are two kinds of transformations of a frame of discernment. One, called *frame refinement*, accounts for learning new facts that make the universe more detailed; another, called *frame coarsening*, accounts for the circumstances that make some of the previous knowledge irrelevant and thus simplify the universe. In terms of sets, refinement partitions some of the sets within a frame of discernment, and the coarsening merges some of the sets together. The analysis starts with the refinement

and then proceeds to the coarsening operation. The definitions are according to [82].

Definition 4.1.1 (Frame refinement) Let $\Theta = \{a_1, a_2, \dots, a_n\}$ be a frame of discernment, let Ω be another frame of discernment. Frame Ω is refined from Θ if for every singleton $\{a_i\} \in 2^\Theta$ there is a subset $\Upsilon_i \in 2^\Omega$, such that $\{\Upsilon_i \subseteq \Omega; i = 1, 2, \dots, n\}$ is a partition of Ω :

1. $\Upsilon_i \neq \emptyset$ for $i = 1, \dots, n$,
2. $\Upsilon_i \cap \Upsilon_j = \emptyset$ for $i \neq j$,
3. $\cup_{i=1}^n \Upsilon_i = \Omega$.

The refinement is a straightforward operation. A single element in the old frame is split into several elements in the new one. It is not difficult to find a function establishing correspondence between these sets. All that is needed is an onto map satisfying a few natural conditions. This function is called a *refining mapping*.

Definition 4.1.2 (Refining mapping) Let Θ and Ω be two frames of discernment and $\sigma : 2^\Theta \rightarrow 2^\Omega$ be a map. σ is a refining mapping if the following conditions are met:

1. The collection $\sigma(\{a_i\})$ forms a partition on Ω ,
2. It is a singleton-union mapping:

$$\sigma : \emptyset \mapsto \emptyset$$

and

$$\sigma : \{a_i\} \mapsto \Upsilon_i = \sigma(\{a_i\})$$

for all singletons in 2^Θ . For non-singletons σ must satisfy

$$\sigma : A = \cup_{a_i \in A} \{a_i\} \mapsto \cup_{a_i \in A} \Upsilon_i.$$

The operation is called a singleton-union mapping because if $\sigma(\{a_i\}) = \Upsilon_i$ then $\sigma(A) = \sigma(\cup_{a_i \in A} \{a_i\}) = \cup_{a_i \in A} \sigma(\{a_i\}) = \cup_{a_i \in A} \Upsilon_i$ for a singleton-union $A = \cup_{a_i \in A} \{a_i\}$. Intuitively, one can think of elements of $\sigma(A) \subseteq \Omega$ as of a detail propositions of proposition A . By the same token, subset $A \subseteq \Theta$ represents a summary proposition of $\sigma(A) \subseteq \Omega$.

Refining a frame of discernment results in a new frame of discernment and, consequently, in a new Kripke model. The new model can always be built from scratch using the procedure introduced in Section 3.3. Depending on the model's complexity, although doing so might be computationally expensive. It is much more effective to have some analogue of map σ on Kripke models. While doing so, it is important to keep in mind that the definition of refining mapping on frames of discernment is independent of possible mass assignments over the refined frames. Clearly, it cannot be the case for the Kripke models. The same frame of discernment may give rise to different Kripke models depending on its mass assignments. So, at least one assumption must be made before proceeding any further: the new partition elements in the refined frame should have non-zero mass assignments. The requirement is formalised in due course. At the moment, let us look at a simple example.

Let Θ and Ω be two frames of discernment, let \mathfrak{F} and \mathfrak{F}' be corresponding Kripke models, and let σ be a map between 2^Θ and 2^Ω . There should be a procedure transforming \mathfrak{F} into \mathfrak{F}' . Given frame of discernment Θ , the corresponding Kripke model $\mathfrak{F} = \langle W, R, V, m_w \rangle$ must have a node for each subset $\theta \in \Theta$, such that $m(\theta) > 0$. This condition is obvious in the context of the semantic interpretation of frames of discernment, but it will later help to determine the representational limits of the interpretation. The relation R is fully determined by the inclusion relation on Θ and by the rules given in Definition 3.3.3. According to the definition of refinement, only the sets corresponding to the terminal nodes are partitioned. In terms of Kripke models, a partitioning results in adding more terminal nodes to W and updating R accordingly.

Using the new partitions as a basis for adding new nodes to the set of possible worlds may increase the complexity of the model and does not always provide the most effective reasoning tool. On the bright side, adding new nodes is much less computationally expensive than building a new Kripke model from scratch. A few examples in the next section should help to see the advantages and drawbacks of this approach.

4.1.1 Refinement example

This section presents an example of a frame refinement based on the principles outlined above. The example demonstrates how different mass assignments over the same frame of discernment result in different Kripke models. To spare the effort needed for inventing yet another faux-detective story Shafer's examples are revisited. Recall the

burglary of the sweetshop example in Section 3.1.2. The frame of discernment after evidence combination is $\Theta = \{LI, LO, RI, RO\}$. Now assume that Sherlock Holmes has some idea about the sex of the burglar in each possible situation.

It is important to note the difference between introducing a new piece of evidence and refining a frame. In case of introducing a new piece of evidence, some beliefs about the burglar being male or female are introduced and then combined with already known information leading to updated beliefs about the burglar's identity. It is not the case with the refinement. The belief in the hypothesis about thief being a left-handed insider is not amended; the already known beliefs are redistributed between female and male left-handed insiders. The belief in the general proposition about the burglar being a left-handed insider stays unaffected. This assumption is in line with the discussion about the role of the mass assignment in forming a Kripke model and is not always true for a general frame of discernment.

Gaining knowledge about the sex of the intruder changes the initial frame Θ into $\Omega = \{lif, lof, lim, lom, rif, rof, rim, rom\}$, the map $\sigma : 2^\Theta \rightarrow 2^\Omega$ is obvious:

$$\begin{aligned}\sigma(\{LI\}) &= \{lif, lim\}, \\ \sigma(\{LO\}) &= \{lof, lom\}, \\ \sigma(\{RI\}) &= \{rif, rim\}, \\ \sigma(\{RO\}) &= \{rof, rom\}.\end{aligned}\tag{4.1}$$

The sets $\Upsilon_i = \{*m, *f\}$, where $*$ = LI, LO, RI, RO , form a partition of Ω . The rest of the conditions from Definition 4.1.2 is satisfied by construction, so σ is a refinement. No specific pieces of evidence $\{M, F\}$ were introduced into the picture. If a belief about the sex of the offender is to be calculated, it has to be retrieved from the appropriate statements.

First, assume that every singleton in the refined frame Ω has a non-zero mass assignment. Let us see why this assumption must be made. Consider map σ defined in equation (4.1). Assume now that one of the singletons in Ω has a mass assignment equal to zero, say $\{lif\}$. But this means that Ω does not discern this proposition. From the semantic representation's point of view the frame of discernment becomes $\Omega = \{lof, lim, lom, rif, rof, rim, rom\}$ and the refining mapping is no longer defined

by equation (4.1). The new map σ' is given below:

$$\begin{aligned}
 \sigma'(\{LI\}) &= \{lim\}, \\
 \sigma'(\{LO\}) &= \{lof, lom\}, \\
 \sigma'(\{RI\}) &= \{rif, rim\}, \\
 \sigma'(\{RO\}) &= \{rof, rom\}.
 \end{aligned} \tag{4.2}$$

Equations (4.1) and (4.2) have different images for set $\{LI\}$ and thus represent two different refining maps. Every singleton in Ω should merit a new node in the corresponding Kripke model, unless there is a condition about assigning new nodes only to the elements with non-zero mass assignments.

Let us now return to the refinement described by σ . Taking a little leap of reasoning rectified later, relation R is extended in the obvious way: $LI = lim \cup lif$. So, because $lif \subset LI$, $lim \subset LI$, and $V(LI) = \{w_8\}$, $V'(lim) = \{w_9\}$, $V'(lif) = \{w_{10}\}$, we have $w_8 R w_9$, $w_8 R w_{10}$ and so on.

The old Kripke frame corresponding to the example is shown in Figure 3.8 on page 58. The new evidence requires every subset represented by a terminal node of the frame to be split in two. So, each terminal node from the old model sees two new nodes. The new model is shown in Figure 4.1. The formulae true at w_5, \dots, w_8 are the same as in the old frame. The refined evidence is reflected on the new terminal nodes w_9, \dots, w_{16} . The new validated formulae are not listed because of space considerations.

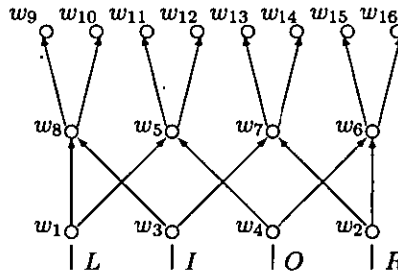


Figure 4.1: Refined burglary of sweetshop I

4.1.2 Refining versus evidence combination

The difference between the frame refinement and the evidence combination becomes clearer after looking at the example below. Let us see the effects of introducing the sex of the offender into the setup. Assume that the evidence about the offender's sex is available based only on the knowledge that the burglar is an insider. The

refined frame of discernment is then $\Omega = \{lif, rif, lim, rim, RO, LO\}$. Such a frame is not a result of combining $\Theta = \{LI, LO, RI, RO\}$ with $\{M, F\}$. Yet it is a perfectly legitimate refinement¹. The only difference is in σ , which, in order to be representable by a Kripke model, becomes

$$\begin{aligned}\sigma(LI) &= (lif, lim), \\ \sigma(RI) &= (rim, rif), \\ \sigma(RO) &= RO, \\ \sigma(LO) &= LO.\end{aligned}$$

The corresponding Kripke frame is shown in Figure 4.2. Combining two independent

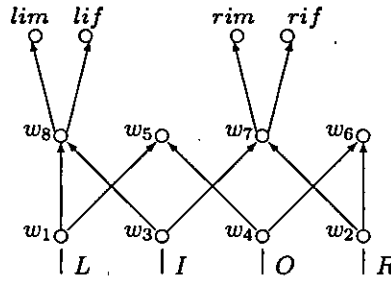


Figure 4.2: Refined burglary of sweetshop II

frames of discernment $\Theta = \{LI, LO, RI, RO\}$ and $\Theta' = \{M, F\}$ according to the procedure in Section 3.3 results in yet another Kripke frame in Figure 4.3. Even though it is possible to construct models with equal mass assignments of the corresponding nodes, the models in Figures 4.1 and 4.3 do not verify the same sets of formulae.

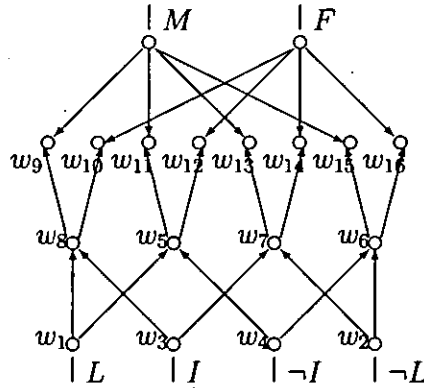


Figure 4.3: burglary of sweetshop after another evidence combination

¹Introducing additional conditions such that $RO \cap M = LO \cap M = LO \cap F = RO \cap F = \emptyset$ may result in $\Omega = \{lif, rif, lim, rim, RO, LO\}$. Such models will not be semantically equivalent.

4.1.3 Formalising model refinement

The example above is very straightforward and logical, but not too strict. The reasoning is as follows: without much justification, a new node was assigned to each singleton of the refined frame. While doing so is not necessarily wrong, it may not always be the case. Even though it is straightforward, the procedure of building the models from frames of discernment first calls for translating sets in a frame of discernment to statements in propositional language \mathcal{L} .

The definition of refinement on frames of discernment does not take the mass assignments into consideration. It is only concerned with the sets discerned by a frame. The Kripke models considered here should also take into account the mass assignments. It only makes sense to assign new nodes to the sets in the core of the refined model's support function. To proceed further this condition must be made explicit.

Definition 4.1.3 (Set W' on refined frames) Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ be a Kripke model over frame of discernment Θ ; and let $\mathfrak{G} = \langle W', R', V', m'_w \rangle$ be its refinement over frame Ω . Let m_Ω be the mass assignment over Ω . Define the set of possible worlds of \mathfrak{G} as

$$W' := W \cup \{w_A : m_\Omega(A) > 0, A \subset \Omega\}. \quad (4.3)$$

Definition 4.1.4 (Relation R' on refined frames) Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ be a Kripke model and $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ its refinement. Relation R' satisfies the following properties

1. $R \subseteq R'$;
2. $w \not R' v, v \not R' w$ for all $v, w \in W' \setminus W$;
3. $w R' v$ and $v \in W' \setminus W$ implies w is a terminal node in \mathfrak{F} ($\{w\} \upharpoonright R = \{w\}$).

The next restriction ensures that the new mass assignments actually *refine* the already known facts rather than just reshuffle the known things randomly. The condition amounts to checking whether the core elements of the new support function are within the partitions of the refined frame.

Proposition 4.1.1 (Conditions on refinement's support function) Let frame Ω be a refinement of frame Θ , let $\sigma : 2^\Theta \rightarrow 2^\Omega$ be the refining mapping, S a support function on Θ , and S' a support function on Ω . Then S' must satisfy the following condition:

$$\{x\} \in C(S') \text{ only if } x \in \sigma(\{y\}) \text{ for some } \{y\} \in C(S), \quad (4.4)$$

where $C(\cdot)$ is a core of a support function.

Proof: Assume the opposite (that $\{y\} \notin C(S)$). This means that support for $\{y\}$ is zero. Therefore, set $\{x\} \subseteq \sigma(\{y\})$ also has zero support by monotonicity property of support function with respect to set inclusion, which is a contradiction. \square

The definition above makes restrictions on the possible mass assignments explicit. The restrictions are needed because the operations do not produce a totally new frame of discernment, but rather modify the existing one. The mass assignment on the refined frame should not contradict any previously gained knowledge. The mass assignment restricted according to (4.4) will only result in a model within the limits of the developed logic.

Consider the burglary of the sweetshop again. Assume that Sherlock Holmes knows something about the sex of the burglar, but not for every possible situation. In other words, assume the same refining mapping σ , but a new mass assignment. The new knowledge is given by the mass assignments below

$$\begin{aligned} m(\{lim\}) &= s_1, \\ m(\{lif\}) &= s_2, \\ m(\{lom, lof, rim\}) &= s_3, \\ m(\{lim, lif, rif\}) &= s_4, \\ m(\text{everywhere else}) &= 0. \end{aligned} \quad (4.5)$$

The mass assignments above give rise to a support function S' , but the conditions of equation (4.4) are violated. For instance, $\{lom, lof, rim\}$ is a core element of S' , but there is no singleton $\{x\}$ in the core of support function S such that $\{lom, lof, rim\} \subseteq \sigma(\{x\})$. Building a refined Kripke model with the mass assignments of the terminal nodes given by equation (4.5) is still possible, but this model would not result from refining the model in Figure 3.8. If the model given in Figure 4.2 is kept, then assigning masses to the terminal nodes of the model becomes quite an impossible task. Following the earlier procedure will result in assigning some mass to a non-terminal node, which leads to contradiction.

Refining frames of discernment has several nice mathematical properties, which should be transferred to Kripke models. To see whether a property is preserved, the formal description of the proposed procedure must be completed. The limitations applied to the support functions and ordering relations were already shown. Below,

it is shown how the sets of possible worlds in two models are linked. The universes under consideration are finite, and thus it is always possible to find the smallest subset with a non-zero basic probability assignment.

Let Θ be a frame of discernment, let Ω be its refinement, let $\sigma : 2^\Theta \rightarrow 2^\Omega$ be the refining map, and $\mathfrak{F} = \langle W, R, V, m_w \rangle$ and $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ be the Kripke models corresponding to Θ and Ω . According to the just outlined procedure $W \subseteq W'$ and $R \subseteq R'$. One can even make a stronger statement about the ordering relations in both frames $X \upharpoonright R \subset X \upharpoonright R', \forall X \in W$.

Consider $w_i \in W$ such that $V(a_i) = \{w_i\}$ and $\sigma(\{a_i\}) = \Upsilon_i$. What happens with the valuation of a_i on Ω ? $V'(\sigma(a_i))$ must include w_i . Because the operation just adds new nodes, the model must also include the nodes corresponding to Υ_i :

$$V'(\sigma(a_i)) = V(a_i) \cup V'(\Upsilon_i) = V(a_i) \cup V'(\cup_j v_j^i) \quad (4.6)$$

$V(a_i)$ is known and $V(\cup_j v_j^i) = \emptyset$, therefore $V'(\cup_j v_j^i)$ must correspond to nodes in $W' \setminus W$. Thus, a new node has to be assigned to every v_j^i whose mass assignment is non-zero $m_\Omega(v_j^i) > 0$ and that $\nexists \zeta \in \Omega$ such that $m_\Omega(\zeta) > 0$ and $\zeta \subset v_j^i$. Υ_i s form a partition of Ω , v_j^i s in their own partition Υ_i , and thus $\cup_i(\cup_j v_j^i) = \Omega$.

4.2 Properties of refinement

Before exploring the properties of the refined frames, it must be established whether the frames in question are well-defined. If the Kripke model analog of the refinement is defined correctly, every refinement mapping will result in a unique Kripke frame.

Proposition 4.2.1 *Kripke model refinement is a well-defined operation.*

Proof. Let $\mathfrak{F}_1 = \langle W_1, R_1, V_1, m_w^1 \rangle$ and $\mathfrak{F}_2 = \langle W_2, R_2, V_2, m_w^2 \rangle$ be two Kripke models corresponding to frame of discernment Λ , both constructed according to definitions 3.3.2- 3.3.4. Two models are identical up to a node permutation if there is a permutation $\pi : W_1 \rightarrow W_2$ such that $\forall u, v \in W_1$ having uR_1v implies $\pi(u)R_2\pi(v)$. The cardinalities of both sets are equal by construction, and thus only one-to-one maps are to be considered. Assume there is no such map. Then, there are nodes $x_1, y_1 \in W_1$, such that $x_1R_1y_1$ but $\pi(x_1) \not R_2 \pi(y_1)$ for all $\pi : W_1 \rightarrow W_2$. According to the procedure for building Kripke models, if there is a node $x_1 \in W_1$ then there is a set $\lambda \subseteq \Lambda$ such that $V_1(\lambda) = x_1$. Therefore $V_2(\lambda)$ is not empty either, so there must be a node $V_2(\lambda) = x_2$. Similarly, if $x_1R_1y_1$ then $\mu \subset \lambda$, where $V_1(\mu) = y_1$, $V_2(\mu) = y_2$ and $x_2R_2y_2$. \square

Thus, every frame of discernment corresponds to a unique, up to a node permutation, Kripke model *before* refinement. Each partition of the frame of discernment produces a unique model, so the refinings and frames of discernment are in one-to-one correspondence, as desired. A well-defined transformation preserves some set operations. The details are in the theorem below quoted from [1].

Theorem 4.2.1 (Preservation properties) *Refining mapping $\sigma : 2^\Theta \rightarrow 2^\Omega$ preserves set operations and relations:*

1. σ is a one-to-one mapping from 2^Θ into 2^Ω , and $\sigma(\Theta) = \Omega$;
2. For all $A, B \subseteq \Theta$

$$\begin{aligned}\sigma(A \cup B) &= \sigma(A) \cup \sigma(B), \\ \sigma(A \cap B) &= \sigma(A) \cap \sigma(B), \\ \sigma(A \setminus B) &= \sigma(A) \setminus \sigma(B), \\ \sigma(\Theta \setminus B) &= \Omega \setminus \sigma(B),\end{aligned}\tag{4.7}$$

Consider a Kripke model $\mathfrak{F} = \langle W, R, V, m_w \rangle$ corresponding to the frame of discernment Θ and the model's refinement $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ corresponding to frame of discernment Ω . Let $\sigma : 2^\Theta \rightarrow 2^\Omega$ be the refining mapping. It is already known that $V'(\sigma(a)) = V(a) \cup V'(\Upsilon)$, where $\Upsilon = \sigma(a)$ for all $a \subseteq \Theta$. The question is whether $V'(\sigma(A \cup B)) = V'(\sigma(A)) \cup V'(\sigma(B))$ or not. Since $\sigma(A \cup B) = \sigma(A) \cup \sigma(B)$, it is enough to know if $V(A \cup B) = V(A) \cup V(B)$ is satisfied. The latter is true according to the Definition 3.2.1 of a valuation map.

The same argument holds for the rest of equalities in equation (4.7). Thus, given the parallelism between the propositions in \mathcal{L} and sets in Θ , the equalities (4.8) directly follow from Definition 3.2.1:

$$\begin{aligned}V(\emptyset) &= \emptyset, \\ V(A \cup B) &= V(A) \cup V(B), \\ V(A \cap B) &= V(A) \cap V(B), \\ V(A \setminus B) &= V(A) \setminus V(B).\end{aligned}\tag{4.8}$$

In a models' terms, a refining is a simple operation adding extra nodes reachable from some of the original model's terminal nodes. Moreover, the properties of a refinement given by Shafer apply to all possible future mass assignment on the refined frame. The proposed translation procedure only deals with a well-behaved subset of mass assignments. The summary of the properties of a Kripke model's refinement is in the proposition below.

Theorem 4.2.2 (Kripke model refinement) *Let Θ, Ω be two frames of discernment, let $\sigma : 2^\Theta \rightarrow 2^\Omega$ be a refinement mapping and $m : 2^\Omega \rightarrow [0, 1]$ be a mass assignment over Ω satisfying the conditions of equation (4.4). Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ be a Kripke model induced by Θ and $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ by Ω . Call \mathfrak{F}' a refinement of \mathfrak{F} . The following must be true for model \mathfrak{F} and its refinement \mathfrak{F}' .*

1.

$$\forall \{w\} \in 2^\Omega \setminus 2^\Theta \text{ if } m(\{w\}) > 0 \exists v \in W' \text{ such that } V'(\{w\}) = \{v\}; \quad (4.9)$$

2. Relation R is a restriction of R' to set W :

$$\forall w \in W \ (w \upharpoonright R) \subseteq (w \upharpoonright R'); \quad (4.10)$$

3. For any singleton $a \in \Theta$, such that $V(a) = \{w\}$ and $\sigma(a) = \Upsilon$

$$V'(\sigma(a)) = w \upharpoonright R'; \quad (4.11)$$

4. On refined frame \mathfrak{F}' for any $A, B \subseteq \Theta$ the equalities below are true

$$V'(\sigma(A \cup B)) = V'(\sigma(A)) \cup V'(\sigma(B)) = (V(A) \upharpoonright R') \cup (V(B) \upharpoonright R'), \quad (4.12)$$

$$V'(\sigma(A \cap B)) = V'(\sigma(A)) \cap V'(\sigma(B)) = (V(A) \upharpoonright R') \cap (V(B) \upharpoonright R'), \quad (4.13)$$

$$V'(\sigma(A \setminus B)) = V'(\sigma(A)) \setminus V'(\sigma(B)) = (V(A) \upharpoonright R') \setminus (V(B) \upharpoonright R'), \quad (4.14)$$

$$V'(\sigma(\Theta \setminus B)) = V'(\Omega) \setminus V'(\sigma(B)) = W' \setminus (V(B) \upharpoonright R'). \quad (4.15)$$

Proof. Equations (4.9), (4.10) and (4.11) are the immediate consequences of the definitions of the corresponding concepts. Equations (4.12)-(4.15) follow from equation (4.8) and Definition 3.2.1 of validity of formulae. \square

The properties listed in Theorem 4.2.2 are useful, because they allow for the manipulation of models almost as easily as frames of discernment. The properties above are true regardless of the mass assignment on the refined frame. The next question is to which extent the refining can be reversed in the frames of discernment, and what are the corresponding effects of this reversal on the Kripke models.

4.3 Coarsening

The opposite of refining is *coarsening*. In the Dempster-Shafer theory, the coarsening is viewed as an operation secondary to refining. This angle of view works well for the taken approach too. A refining map σ is not necessarily a bijection. The existence of the inverse, σ^{-1} , is then not guaranteed, but some approximation can be produced. This approximation is the coarsening map. The coarsening comes in two different flavors: *inner* and *outer coarsening*.

Definition 4.3.1 (Coarsening) Assume that $\sigma : 2^\Theta \rightarrow 2^\Omega$ is a refining. Mapping

$$\sigma^{-1-} : 2^\Omega \rightarrow 2^\Theta$$

is then called the *inner coarsening* of σ . It is defined by

$$\sigma^{-1-}(\Upsilon) = \{x \in \Theta \mid \sigma(\{x\}) \subseteq \Upsilon\}.$$

Mapping

$$\sigma^{-1+} : 2^\Omega \rightarrow 2^\Theta$$

is called the *outer coarsening* of σ and is defined by

$$\sigma^{-1+}(\Upsilon) = \{x \in \Theta \mid \sigma(\{x\}) \cap \Upsilon \neq \emptyset\}$$

for all $\Upsilon \subseteq \Omega$.

Given the nice properties of the refinement, there is some expectation of convenient algebraic properties from its converse. Just as expected, some set operations are preserved, but the set-preserving properties of coarsening are much weaker than for refining.

Theorem 4.3.1 (Coarsening properties, Shafer 1974) *The inner and outer coarsenings preserve set operations and the set relation:*

1. σ^{-1+} preserves set union:

$$\sigma^{-1+}(P \cup Q) = (\sigma^{-1+}(P)) \cup (\sigma^{-1+}(Q)), \quad (4.16)$$

for all $P, Q \subseteq \Omega$;

2. σ^{-1-} preserves set intersection:

$$\sigma^{-1-}(P \cap Q) = (\sigma^{-1-}(P)) \cap (\sigma^{-1-}(Q)), \quad (4.17)$$

for all $P, Q \subseteq \Omega$;

3. For set difference, the following is true for all $R \subseteq \Omega$:

$$\begin{aligned} \sigma^{-1+}(\Omega \setminus R) &= \Theta \setminus \sigma^{-1-}(R), \\ \sigma^{-1-}(\Omega \setminus R) &= \Theta \setminus \sigma^{-1+}(R); \end{aligned} \quad (4.18)$$

4. σ^{-1-} and σ^{-1+} preserve set inclusion: if $P \subseteq Q$ in 2^Ω then

$$\begin{aligned} \sigma^{-1-}(P) &\subseteq \sigma^{-1-}(Q), \\ \sigma^{-1+}(P) &\subseteq \sigma^{-1+}(Q). \end{aligned} \quad (4.19)$$

The immediate consequence of Theorem 4.3.1 fills in the gaps in the inner/outer coarsening — union/intersection paradigm:

Theorem 4.3.2 (Coarsening and set operations) *Let the setup be the same as in the previous theorem. Then the following are true:*

$$\sigma^{-1-}(P \cup Q) \supseteq \sigma^{-1-}(P) \cup \sigma^{-1-}(Q) \quad (4.20)$$

and

$$\sigma^{-1+}(P \cap Q) \subseteq \sigma^{-1+}(P) \cap \sigma^{-1+}(Q) \quad (4.21)$$

for all $P, Q \subseteq \Omega$ [1].

Clearly, $\sigma^{-1+} = \sigma^{-1-}$, when σ is an isomorphism. If it is not, the relationship is not as straightforward, but still quite predictable. The details are in the theorem below.

Theorem 4.3.3 *Let $\sigma: 2^\Theta \rightarrow 2^\Omega$ be a refining map, let σ^{-1-} and σ^{-1+} be inner and outer coarsenings, then the following equalities are true:*

1. $\sigma^{-1-}(\emptyset) = \emptyset = \sigma^{-1+}(\emptyset)$;
2. $\sigma^{-1-}(\Omega) = \Theta = \sigma^{-1+}(\Omega)$;
3. $\sigma^{-1-}(\Upsilon) \subseteq \sigma^{-1+}(\Upsilon)$ for all $\Upsilon \subseteq \Omega$;

4. $\sigma^{-1-}(\sigma(A)) = A = \sigma^{-1+}(\sigma(A))$ for all $A \subseteq \Theta$, i.e. $\sigma^{-1-}\sigma = \sigma^{-1+}\sigma = \text{id}_\Theta$;
5. $\sigma(\sigma^{-1-}(\Upsilon)) \subseteq \Upsilon$ and $\Upsilon \subseteq \sigma(\sigma^{-1+}(\Upsilon))$ for all $\Upsilon \subseteq \Omega$;
6. $\sigma(A) \subseteq \Upsilon$ iff $A \subseteq \sigma^{-1-}(\Upsilon)$ and $\Upsilon \subseteq \sigma(A)$ iff $\sigma^{-1+}(\Upsilon) \subseteq A$ [82].

Whenever a coarsening is applied to a frame of discernment, the effect on the corresponding Kripke model amounts to reducing the number of nodes and updating the ordering relation accordingly. The only cumbersome parts of the procedure are updating the masses of the nodes and ensuring that the new mass assignments still give rise to a mass function. Below, these effects are described in some detail and the analogies are drawn between the coarsening and its Kripke model counterpart. Just as it happened with the refinement, the scope of the procedure in the realm of Kripke models is narrower. An illustrative example opens the exposition followed by the rigorous part.

4.3.1 Coarsening and Kripke models

Assume that $\mathfrak{F} = \langle W, R, V, m_w \rangle$ is a Kripke model on frame of discernment Θ , let $w_1, w_2 \in W$ be two nodes in the model. A fragment of \mathfrak{F} is shown in Figure 4.4(a). Now, assume that $\sigma : 2^\Theta \rightarrow 2^\Omega$ is a refining map transforming Θ into Ω . $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ is a Kripke model on Ω , where $V'(\sigma(w_1)) = \{w_1, w_3, w_4\}$ and $V'(\sigma(w_2)) = \{w_2, w_5, w_6\}$. The relevant fragment of \mathfrak{F}' is given in Figure 4.4(b). Let

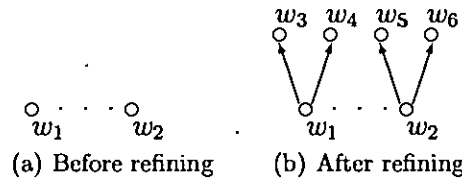


Figure 4.4: Frame and its refinement

$A, B \subseteq \Theta$ be subsets satisfying $V(A) = \{w_1\}$ and $V(B) = \{w_2\}$. Now if $\sigma(A) = \Upsilon_A$ and $\sigma(B) = \Upsilon_B$ in model \mathfrak{F}' , valuations are given by $V'(\Upsilon_A) = \{w_1, w_3, w_4\}$ and $V'(\Upsilon_B) = \{w_2, w_5, w_6\}$. Coarsening has the predictable result of removing some of the terminal nodes: $V(\sigma^{-1-}(\Upsilon_A)) = V(\sigma^{-1+}(\Upsilon_A)) = V(A) = \{w_1\}$ according to Theorem 4.3.3. In this case, the inner and outer coarsenings are equal. It happened because the coarsened sets were the images of singletons in the original frame. If the coarsened sets were not images of singletons in the original, then the results of inner and outer coarsening might differ. To see the difference, consider set Υ_C such that

$V'(\Upsilon_C) = \{w_3, w_4, w_5\}$. Now $V(\sigma^{-1-}(\Upsilon_C)) = \{w_1\}$, but $V(\sigma^{-1+}(\Upsilon_C)) = \{w_1, w_2\}$. The coarsening is a 'reversal operation' on frames of discernment which can only be defined in terms of *previously conducted* refining. As the example above shows, it is different on Kripke models: by definition, the nodes corresponding to the refining partitions are the terminal nodes, and the coarsening of a Kripke model results in removing some of its terminal nodes. To understand the effects of removing the terminal nodes on the truth values of different propositions from \mathcal{L} , some interpretation of V' is needed.

The examples above demonstrated that in the models the coarsening amounted to removing some terminal nodes from the model and updating the ordering relation on the set of possible worlds accordingly. Removing all the edges in the downward closure of the removed nodes reflects outer coarsening, while removing only the edges stemming from the nodes whose upward closure is within the coarsened set reflects inner coarsening.

Definition 4.3.2 (Coarsening on set of nodes) Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ and $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ be two Kripke models such that $W \subset W'$ and $R = R' \cap (W \times W)$. Define outer coarsening $\rho^{-1-} : 2^{W'} \rightarrow 2^W$ and inner coarsening $\rho^{-1+} : 2^{W'} \rightarrow 2^W$, such that

$$\rho^{-1-}(X) = \{w \in W \mid (w \upharpoonright R') \setminus \{w\} \subseteq X\}, \quad (4.22)$$

$$\rho^{-1+}(X) = \{w \in W \mid w \in X \upharpoonright R', (w \upharpoonright R') \cap X \neq \emptyset\}, \quad (4.23)$$

where $X \in \text{Up}W'$ and $X \cap W = \emptyset$.

The condition $X \cap W = \emptyset$ ensures that sets X only contain newly added terminal nodes. If this condition is not met, then X cannot correspond to any set Υ forming a partition and cannot be coarsened. Condition $wRy \Rightarrow y \in X$ in equation (4.23) cancels the transitivity of R' : only the immediate predecessors of the terminal nodes should be present in the coarsened subset. The same condition in equation (4.22) is guaranteed by $(w \upharpoonright R') \setminus \{w\} \subseteq X$. The relationship between valuation functions of both models follows directly from Definition 4.3.2. Both the inner and the outer coarsenings are maps between 2^Ω and 2^Θ , and maps ρ^{-1-} and ρ^{-1+} link $2^{W'}$ and 2^W . Maps $V : \text{Var}\mathcal{L} \rightarrow 2^W$ and $V' : \text{Var}\mathcal{L} \rightarrow 2^{W'}$ are the maps between elements of frames of discernment and sets of possible worlds. The coarsening of a frame of

discernment can thus be related to the coarsening of a Kripke model through

$$\begin{aligned} V(\sigma^{-1-}(X)) &= \rho^{-1-}(V'(X)) \\ V(\sigma^{-1+}(X)) &= \rho^{-1+}(V'(X)) \text{ for any } X \in \Omega. \end{aligned} \quad (4.24)$$

The left hand sides of the equations in (4.24) have V , and the right hand sides have V' : the result of changing the domain. Left hand sides describe the facts in model \mathfrak{F} , right hand sides tell us about \mathfrak{F}' .

4.3.2 Coarsening example

Let $\Theta = \{a_1, a_2, a_3\}$ be a frame of discernment, let $\Omega = \{r_{11}, r_{21}, r_{22}, r_{31}, r_{32}, r_{33}\}$ be its refining. The refining mapping is then $\sigma(\{a_i\}) = \{r_{ij} \in \Omega\}$, and it is a one-to-one correspondence between 2^Θ and 2^Ω . Let us now consider coarsening different sets in Ω . In the simplest case, the results of a refining are simply reversed $\sigma^{-1-}(\sigma(A)) = \sigma^{-1+}(\sigma(A))$, the equality holds for any $A \subseteq \Theta$. Consider now $X \subseteq \Omega$: which is not a result of refining any set in Θ : $\nexists A \subseteq \Theta$, such that $\sigma(A) = X$. The inner and outer coarsening of this set are no longer equal. For example, take $X = (\{r_{11}, r_{21}\})$, there is no set $A \subseteq \Theta$ such that $\sigma(A) = X$, therefore the inner and outer coarsenings do not need to be equal: $\sigma^{-1-}(X) = \sigma^{-1-}(\{r_{11}, r_{21}\}) = \{a_1\}$, but $\sigma^{-1+}(X) = \sigma^{-1+}(\{r_{11}, r_{21}\}) = \{a_1, a_2\}$.

Let us now translate the example in the previous paragraph to Kripke frames. Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ be the frame representing Θ , and let $\mathfrak{G} = \langle W', R', V', m'_w \rangle$ represent frame of discernment Ω . The models satisfy the conditions of the definition of a frame refinement, so \mathfrak{G} is a refinement of frame \mathfrak{F} .

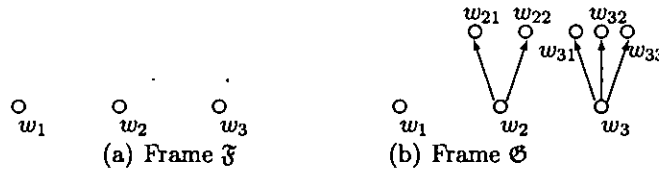


Figure 4.5: Coarsening example

Equations (4.23) and (4.22) are used to calculate the coarsenings of different subsets in W' . Take $A = \{w_1, w_2\} \subseteq W$. Indeed, if $V(a_2) = \{w_2\}$, where $a_2 \in \text{Var}\mathcal{L}$, then $V'(\sigma(a_2)) = A$, so inner and outer coarsenings of A should be equal: $\rho^{-1-}(A) = \rho^{-1+}(A) = \{w_2\}$. Consider now $X = \{w_{22}, w_{31}\} \subseteq W'$. There is no formula ϕ , such that $V'(\sigma(\phi)) = X$, so the inner and outer coarsenings of the set should not be equal:

$\rho^{-1-}(X) = \emptyset$, but $\rho^{-1+}(X) = \{w_2, w_3\}$. The calculations can be continued in a similar fashion for the rest of the subsets in 2^Ω . The natural question is: if m_Θ and $m_{\mathfrak{F}}$ are the mass assignments over frames \mathfrak{F} and Θ , would the results of coarsening mean that in case of outer coarsening $m_{\mathfrak{F}}(\{w_2, w_3\}) = m_{\mathfrak{F}}(\{w_{22}\}) + m_{\mathfrak{F}}(\{w_{31}\})$? Even though the intuition tells us that it should be the case, giving the answer requires further analysis.

In the examples above, the issues of mass assignments in the coarsened frames were deliberately left out. The issue of coarsening the mass assignments is quite far from trivial in both frames of discernment and Kripke models. Moreover, coarsening plays a pivotal role in classifying the support functions. The discussion on the topic involves a few references to the formal properties of the operations in question, namely, coarsening. So, to proceed further some time has to be devoted to the formal properties of a frame coarsening.

4.3.3 Properties of coarsening Kripke models

The theorem below gives a little insight into the relationship between the coarsenings in frame of discernments and Kripke model universes. Most of the properties are quite predictable and almost immediately follow from the relevant definitions and corresponding properties of the frames of discernment.

Theorem 4.3.4 (Properties of coarsening) *Let $\sigma : 2^\Theta \rightarrow 2^\Omega$, be a refining of Θ into Ω , σ^{-1-} be an inner and σ^{-1+} an outer coarsenings of Ω . Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ and $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ be the corresponding Kripke models, $\rho^{-1-}, \rho^{-1+} : 2^{W'} \rightarrow 2^W$ be \mathfrak{F}' inner and outer coarsenings. Then for any sets $P, Q \subset \Omega$ the following are true:*

$$V(\sigma^{-1+}(P \cup Q)) = \rho^{-1+}(V'(P)) \cup \rho^{-1+}(V'(Q)) \quad (4.25)$$

$$V(\sigma^{-1-}(P \cap Q)) = \rho^{-1-}(V'(P)) \cap \rho^{-1-}(V'(Q)) \quad (4.26)$$

$$V(\sigma^{-1-}(\Omega \setminus P)) = W \setminus \rho^{-1+}(V'(P)) \quad (4.27)$$

$$V(\sigma^{-1+}(\Omega \setminus P)) = W \setminus \rho^{-1-}(V'(P)) \quad (4.28)$$

$$P \subseteq Q \Rightarrow \rho^{-1-}(V'(P)) \subseteq \rho^{-1-}(V'(Q)) \quad (4.29)$$

$$P \subseteq Q \Rightarrow \rho^{-1+}(V'(P)) \subseteq \rho^{-1+}(V'(Q)) \quad (4.30)$$

$$V(\sigma^{-1+}(P \cup Q)) \supseteq \rho^{-1-}(V'(P)) \cup \rho^{-1-}(V'(Q)) \quad (4.31)$$

$$V(\sigma^{-1-}(P \cap Q)) \subseteq \rho^{-1+}(V'(P)) \cap \rho^{-1+}(V'(Q)) \quad (4.32)$$

$$\rho^{-1+}(\emptyset) = \rho^{-1-}(\emptyset) = \emptyset \quad (4.33)$$

$$\rho^{-1+}(V'(\Omega)) = \rho^{-1-}(V'(\Omega)) = W \quad (4.34)$$

$$\rho^{-1-}(X) \subseteq \rho^{-1+}(X) \forall X \subseteq W'. \quad (4.35)$$

Proof sketch. The proof is based on putting together definitions of the operations in question. To demonstrate the line of reasoning equation (4.25) is proved.

$V(\sigma^{-1+}(P \cup Q)) = V(\sigma^{-1+}(P) \cup \sigma^{-1+}(Q))$ follows from equation (4.16), according to properties given in equation (4.8) $V(\sigma^{-1+}(P) \cup \sigma^{-1+}(Q)) = V(\sigma^{-1+}(P)) \cup V(\sigma^{-1+}(Q))$, which after substitution given in equation (4.16) becomes $V(\sigma^{-1+}(P)) \cup V(\sigma^{-1+}(Q)) = \rho^{-1+}(V'(P)) \cup \rho^{-1+}(V'(Q))$ as desired.

The rest of identities in Theorem 4.3.1 are proved using the argument very similar to the one shown above. \square

Equations (4.27) and (4.28) are the immediate consequences of equalities (4.18). The only point to be checked is whether $V(\Theta) = W$. Surprisingly, this equality did not appear in the discussion earlier. According to Definition 3.3.3 there is a node for every singleton $a \in \Theta$, such that $m(a) > 0$. Thus, $V(\Theta)$ is the set of nodes where being a part of any subset of Θ is true, which includes all the nodes of the model, or the whole set W .

There is no immediate analogue to this property on frames of discernment. The valuation of the core of a support function on a frame always equals the totality of the possible worlds. On the other hand, the core of a support function is not necessarily equal to the frame's universal set. Speaking strictly, the semantic models did not represent abstract frames, they represented *particular* belief functions over the frames. In the next section, this relationship is addressed in more detail.

4.3.4 Mass assignment and updating frames

The redistribution of the masses in case of a frame refinement was straightforward in both frames of discernment and Kripke models. The frame refinement redistributes the belief already assigned to a set among its non-overlapping subsets. So, to update the corresponding support functions the masses of the old focal elements should be redistributed among the new ones. Such transformation was just as easy in the case of Kripke models, the mass assignment of old terminal nodes was pushed to new terminal nodes they could reach.

To illustrate the point of the previous paragraph, let us again look at Figures 4.5(a) and 4.5(b). Assume that in model \mathfrak{F} , prior to the refinement, we have masses

$m_{\mathfrak{F}}(w_1)$, $m_{\mathfrak{F}}(w_2)$ and $m_{\mathfrak{F}}(w_3)$ assigned to corresponding nodes. Denote the mass assignments on refined frame \mathfrak{G} by $m_{\mathfrak{G}}$. Take $m_{\mathfrak{F}}(w_2)$. After the refinement this mass should be distributed between the new terminal nodes w_{21} and w_{22} . The equality have $m_{\mathfrak{G}}(w_{21}) + m_{\mathfrak{G}}(w_{22}) = m_{\mathfrak{F}}(w_2)$ should also hold.

Coarsening the sets of terminal nodes accessible from a single non-terminal node does not present much trouble. Let the masses of nodes w_{31} , w_{32} and w_{33} be given by $m_{\mathfrak{G}}(w_{31})$, $m_{\mathfrak{G}}(w_{32})$ and $m_{\mathfrak{G}}(w_{33})$. Coarsening $X = \{w_{31}, w_{32}, w_{33}\}$ would not pose any problem. Results of both inner and outer coarsening are w_3 ; the mass assignment on the coarsened model should be $m_{\mathfrak{F}}(w_3) = m_{\mathfrak{G}}(w_{31}) + m_{\mathfrak{G}}(w_{32}) + m_{\mathfrak{G}}(w_{33})$.

In a set whose coarsening is less trivial, the situation becomes slightly more complicated. Let $Y = \{w_{21}, w_{22}, w_{31}, w_{32}\}$, then $\rho^{-1-}(Y) = w_2$, but the mass assignment $m_{\mathfrak{F}}(w_2) = m_{\mathfrak{G}}(w_{21}) + m_{\mathfrak{G}}(w_{22}) + m_{\mathfrak{G}}(w_{31}) + m_{\mathfrak{G}}(w_{32})$ is no longer equal to its pre-refinement mass. Outer coarsening becomes $\rho^{-1+}(Y) = \{w_2, w_3\}$. The mass assignment now becomes even trickier: some portion of it should be assigned to node w_2 and the remainder should be assigned to node w_3 , which is not always a terminal node. It is not clear what happened with the rest of the terminal nodes they could reach. The only reasonable, if excessive, solution to introduce a new node w_r which 'stores' the leftover portion of coarsened mass. First, define a set

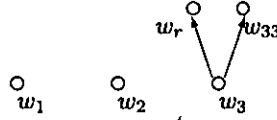


Figure 4.6: Frame \mathfrak{G} after coarsening

$Y^* = \{w \in Y \mid \exists x \in v \mid R' \setminus \{v, w\}, vR'w, x \notin Y\}$. The mass of this node is given by $m_{\mathfrak{F}}(w_r) = \sum_{w \in W', w \in Y^*} m_{\mathfrak{G}}(w)$. The expressions in the previous phrase are awkward, but have a very simple meaning. Set Y^* is a collection of nodes in Y that belong to some upward closure not fully contained in Y . The upward closure of node w_2 is fully contained in Y , but the upward closure of w_3 is not, so $Y^* = \{w_{31}, w_{32}\}$. The mass assignment node w_r (r for remainder) is just a sum of masses in Y^* . In the example under scrutiny it is $m(w_r) = m_{\mathfrak{G}}(w_{31}) + m_{\mathfrak{G}}(w_{32})$. Frame \mathfrak{G} then takes shape shown in Figure 4.6.

There is a problem though: the coarsening operation should be defined for all the sets in $2^{W'}$, so the 'remainder' node in each case will be different. Not using a remainder node solves the problem. Instead, every node in the coarsened frame must have the mass equal to the sum of masses of the terminal nodes it sees. The support for coarsened images of subsets of $2^{W'}$ is then the sum of the masses of new terminal

nodes in it. The definition below gives a formal summary of the observations given in this paragraph.

Definition 4.3.3 (Mass assignment on a coarsened Kripke frame)

Let $\mathfrak{F} = \langle W, R \rangle$, $\mathfrak{G} = \langle W', R' \rangle$ be two Kripke frames, let $\rho^{-1-}, \rho^{-1+} : 2^{W'} \rightarrow 2^W$ be inner and outer coarsening of \mathfrak{G} into \mathfrak{F} . Let $m_{\mathfrak{F}}$, $m_{\mathfrak{G}}$ be the mass assignments over the frames. The mass assignments are then related through

$$m_{\mathfrak{F}}(w) = \frac{1}{N} \sum_{v_i \in W', wR'v_i} m_{\mathfrak{G}}(v_i), \text{ where } w \text{ is a terminal node in } \mathfrak{F}, \quad (4.36)$$

where $N = \sum_{w \in W} (\sum_{wR'v_i} m_{\mathfrak{G}}(v_i))$ is the normalising coefficient. The support for coarsened images of subsets on $2^{W'}$ is calculated by

$$S(\rho^{-1-}(X)) = \sum_{w_i \in \rho^{-1-}(X)} m_{\mathfrak{F}}(w_i), \quad (4.37)$$

or

$$S(\rho^{-1+}(X)) = \sum_{w_i \in \rho^{-1+}(X)} m_{\mathfrak{F}}(w_i), \quad (4.38)$$

where $S : 2^W \rightarrow [0, 1]$ is a belief function over \mathfrak{F} .

The normalising coefficient N is one, when frame \mathfrak{G} is refined from some other frame. N may differ from one if the frame to which coarsening is applied is not a result of a refining. Checking whether mass assignment $m_{\mathfrak{F}}$ gives rise to a belief function is trivial: no new mass assignments are introduced. Instead, the already known assignments are reshuffled to create a different belief distribution, and no further check is needed.

To see how the definition works, let us refer to the same model. Consider the inner coarsening $\rho^{-1-}(\{w_{31}, w_{32}\}) = w_3$, $\rho^{-1-}(\{w_{31}, w_{32}, w_{33}\}) = w_3$, and $\rho^{-1-}(\{w_{31}\}) = w_3$. Continuing the line of reasoning suggested in the previous paragraph and introducing extra nodes storing the leftover beliefs results in the model identical to the original one. On the other hand, it is obvious that the coarsening should not be trivial. A trivial coarsening distinguishes the same propositions with the same mass assignments before and after coarsening.

To substantiate the point, a few coarsened mass assignments are calculated for the model in question. There are six terminal nodes in frame \mathfrak{G} , so set $2^{W'}$ has $2^6 - 1 = 63$ non-empty subsets. Calculating all the mass assignments is not so interesting, and

the first five give a clear idea of the whole process.

$$\rho^{-1-}(\{w_1\}) = \rho^{-1-}(\{w_1, w_{21}\}) = w_1;$$

$$\rho^{-1-}(\{w_{21}\}) = \rho^{-1-}(\{w_{21}, w_{22}\}) = \rho^{-1-}(\{w_{21}, w_{22}, w_{31}\}) = w_2$$

the list could be continued for the rest of the subsets.

The attention now can be directed towards the mass assignments of the nodes in the coarsened frame. The masses are assigned according to the rules given in the original description of the procedure. Recall that only the terminal nodes have non-zero mass assignments. The non-terminal nodes are supported by the evidence visible from them: the support of a non-terminal node is calculated by adding the masses of the terminal nodes accessible from it. According to this procedure $m_{\mathfrak{F}}(w_1) = m_{\mathfrak{G}}(w_1)$ because these nodes are terminal in both models. $m_{\mathfrak{F}}(w_2) = m_{\mathfrak{G}}(w_{21}) + m_{\mathfrak{G}}(w_{22})$ and $m_{\mathfrak{F}}(w_3) = m_{\mathfrak{G}}(w_{31}) + m_{\mathfrak{G}}(w_{32}) + m_{\mathfrak{G}}(w_{33})$. The support attributed to the coarsened images of different subsets of W' is then calculated from the masses of the nodes in their coarsened images. For example, since $\rho^{-1+}(\{w_{22}, w_{31}\}) = \{w_2, w_3\}$, the coarsened support for the propositions verified at w_{22} and w_{31} is equal to $m_{\mathfrak{F}}(w_2) + m_{\mathfrak{F}}(w_3)$ and so on. There are many more different statements in the coarsened frame with same degree of support as there were in the original frame. The coarsened frame does not distinguish between some of the propositions distinct in the original frame. In some sense, the proposed mass assignment approach ensures that the coarsening actually makes the frame cruder.

4.4 Refining, coarsening and support functions

The operations of frame refinement and coarsening serve two fundamental purposes. The first task is to enable the decision maker to implement the new knowledge into an already existing framework. The second purpose is less obvious, but no less important. By analysing the possible refinings and coarsenings on frames of discernment, Shafer was able to prove several important facts about different types of the belief functions. In particular, he proved that every inclusion in the scheme of different support functions shown in Figure 3.1 is a proper inclusion.

According to Shafer, actual evidence can be represented by simple support, separable support and support functions, but not by general belief functions. In the realm of Kripke models, understanding which types of support functions can be expressed through the means of the proposed procedure helps determine the limits of

the approach's applicability. The limits of the interpretation's expressive power were already established. Understanding which types of support functions can be interpreted via the proposed procedure will enable us to understand if the semantics of Int represents the whole of the Dempster-Shafer theory or not.

The answer to the latter question is positive, but with reservations. To understand the reservations, a short overview of Shafer's findings about the support functions is given, and then the results relevant for the approach are chosen.

The work so far operated with the notions of simple and separable support functions and a belief function, but not with a support function. Recall, that the distinction between these functions was discussed in Section 3.1.2. Introducing such a basic notion was postponed until late in the progress because a support function is defined as a coarsening of a belief function and thus could not be introduced any time earlier. The definition below is due to Shafer.

Definition 4.4.1 (Consistent belief functions) *Let Θ and Ω be two frames of discernment, let $\sigma : 2^\Theta \rightarrow 2^\Omega$ be a refining mapping. $Bel_0 : 2^\Theta \rightarrow [0, 1]$ and $Bel' : 2^\Omega \rightarrow [0, 1]$ are belief functions on Θ and Ω . Bel and Bel_0 are consistent if for any set $A \subset \Theta$ the following is true*

$$Bel_0(A) = Bel(\sigma(A)).$$

In this case, Bel_0 is a restriction of Bel to Θ . The restriction of a belief function to a coarsened frame is denoted $Bel_0 = Bel|_{2^\Theta}$.

Whenever a refining is performed according to the proposed procedure on a Kripke model, the original support function is the restriction of the refined one. This observation is not necessarily true for the frames of discernment, but always works for the Kripke models. A Kripke model with mass assignment does not represent a frame of discernment, but a frame of discernment with a belief function over it. In other words, different belief functions over the same frame of discernment may result in different Kripke models. The distinction between support functions and separable support functions is given through the possibility of coarsening a belief function into a separable support function.

Definition 4.4.2 (Support functions) *Let Θ be a frame of discernment. A belief function $Bel : 2^\Theta \rightarrow [0, 1]$ is a support function if there is a refinement Ω of Θ and some separable support function $S : 2^\Omega \rightarrow [0, 1]$, such that $Bel = S|_{2^\Theta}$.*

For every support function, there is a common refinement of a such function with some separable support function. The condition is nothing but demanding some 'middle ground' between support functions and their simpler counterparts which can be reached from both domains. The condition, as it is stated in the definition, is not too useful for determining whether a belief function is a support function or not. The theorem below gives the necessary practical tools.

Theorem 4.4.1 (Conditions on Support functions) *Let Θ be a frame of discernment, let $Bel : 2^\Theta \rightarrow [0, 1]$ be a belief function on Θ , let $C(Bel)$ be its core. The following are equivalent:*

1. *Bel is a support function;*
2. *$C(Bel)$ has a positive commonality number;*
3. *$C(Bel)$ has a positive basic probability assignment [1].*

Theorem 4.4.1 is useful for checking if the belief function belongs to a support functions class. Can Kripke models represent support functions? Can Kripke models represent belief functions that are not support functions? To answer both questions, the examples of belief and support functions are given first. Plenty of simple and separable support functions on Kripke models were already constructed, so the discussion will be centred around support functions and belief functions.

Consider Shafer's example of a belief function that is not a support function. Let $\Theta = \{a, b, c, d\}$, the focal elements of belief function S are $m(\{a, b\}) = s_1$, $m(\{c\}) = 1 - s_1$ and $m(A) = 0$ for all other sets. The core of this belief function is $C(S) = \{a, b, c\}$. According to the last condition of Theorem 4.4.1 this belief function is not a support function because $m(C(S)) = 0$.

Frame Θ with this belief function cannot be represented by a Kripke model, but the reason is different. First of all, there is a singleton in Θ with a zero mass assignment which does not merit a node. Following the developed procedure results in the model no different from the model over a smaller frame of discernment not including $\{d\}$. Thus, the procedure is only capable of representing the beliefs S such that $C(S) = \Theta$. Still, the cases in which the mass assignment of the core was zero were not ruled out.

The distinction between representing a belief function that is not a support function and a support function is somehow trickier. All the belief functions given by the mass assignments in the previous paragraph can be easily represented by two isolated nodes, and in this case the semantic model would not look any different from a model

representing a simple support function centred around a single focal element. The difference is in the semantics of the nodes. While in the case of a simple support function one of the nodes supports some particular piece of evidence, and the other node collects unassigned beliefs, in the case of representing a belief function that is not a support function every node supports some particular piece of evidence. A simple procedure helps to check whether a model represents a belief or a support function.

Recall the procedure used to build a model representing a support function in which some core elements include others. The procedure was described on page 53. The idea of making a copy of each node representing a core element and then updating the relation R accordingly works for the problem at hand too. Let $\mathfrak{F} = \langle W, R, V, m_w \rangle$ be a model, $\mathfrak{F}' = \langle W', R', V', m'_w \rangle$ its refinement, and $\sigma : 2^W \rightarrow 2^{W'}$ a refinement mapping. Define σ as $\sigma(\{w\}) = \{w'\}$ for all terminal nodes w in \mathfrak{F} . Update R' according to the rules. Model \mathfrak{F} represents a support function if there is a node v' in W' , such that $v' \mid R' = \{v'\} \cup W' \setminus W$. If there is no such node, then \mathfrak{F} represents a belief function that is not a support function. The condition can be restated as checking whether applying σ to a model yields a node whose immediate successors' masses amount to one.

The testing procedure outlined in the previous paragraph is in agreement with Shafer's observations. Combining a belief function with a support function produces another belief function, which is in agreement with the earlier observations. The limits of the proposed representation can be outlined and some conclusions about the semantic of both the Dempster-Shafer evidence combination rule and evidence update can be made.

The procedure in Chapter 3 is capable of representing all types of belief functions: belief functions, support functions, separable support functions, and simple support functions. The procedure is only suitable for support functions $S : 2^\Theta \rightarrow [0, 1]$ such that $\mathcal{C}(S) = \Theta$. This limitation is important from the semantic point of view. Consider two frames of discernment Θ and Ω , such that $\Theta \subset \Omega$. Let m_Θ and m_Ω be the mass assignments over them. Define $m_\Omega(X) = m_\Theta(X)$ for all $X \subseteq \Theta$ and let $m_\Omega(Y) = 0$ otherwise. From the evidential point of view these two frames are different, and refining them can yield totally different results. From the semantic point of view the two are identical and will be represented by the same model. Moreover, refinements or coarsenings of the corresponding Kripke models will not represent all the possible results of refinements and coarsenings of Θ and Ω .

The limits of representation are well in line with the original premises outlined in Chapter 2. The logic should not only represent the Dempster-Shafer theory world

view, it should also contribute to building a system for consistent reasoning. The consistency in Kripke models requires that a proposition originally known to be true cannot become false later. The limitation on the mass assignments introduced in this section and on the models themselves makes this change of validity impossible. The requirement for the core of support function to be equal to the frame of discernment was needed to limit the scope only to formulae with non-zero support.

It was already mentioned that Shafer limited actual observable evidence to the situations described by support functions, but not by general belief functions. Exposition above suggests that only the support functions whose core is equal to the frame of discernment are suitable for a consistent inferential apparatus. The main strength of the proposed approach lies in its analogue to the evidence combination rule. Any two Kripke models may be combined according to the analogue of the Dempster-Shafer evidence combination rule. The possibilities for evidence updates done with the aid of frame refinement or coarsening are limited to the transformations which do not violate the semantic integrity of the model: the transformations do not change the set of valid variables in the non-transformed nodes.

}

Chapter 5

Discussion

5

The results presented in this work belong to more than one field: the proposed interpretation of the Dempster-Shafer theory is useful for uncertainty representation; the discussion in Chapter 2 attempts to answer some epistemological questions, the resulting formalism is within limits of multivalued logic. Below, the main contributions are recaptured, and possible future research directions are outlined in a structured way.

5.1 The contribution

The main inspiration behind this work was to construct a logic specifically for the needs of the Dempster-Shafer theory. The apparatus should not only allow for translating the evidential setup to some propositional language, but should also help understand the semantics of the theory. The latter goal differentiates the proposed interpretation from the earlier logical approaches to the Dempster-Shafer theory. Most of the earlier logic interpretations started with analysis of the meaning of the logical connectives, and the functions describing these connectives were chosen early in the development. The approach in this work is different. First, an effort to understand the nature of the mathematical objects viewed through the prism of the Dempster-Shafer theory was made. There are infinitely many propositional languages, and most of them can be used for describing the Dempster-Shafer theory universe. The problem is interesting because there is no objective criterion that governs the choice of a propositional language. When viewed from this angle, the interpretations mentioned above can be classified as being *ad hoc* or naïve. However, the semantics of such logic systems are influenced by the choice of the propositional language. The Dempster-Shafer theory is a universal theory applicable to any kinds of sets and

universes. Having some particular logic connectives for its interpretation is unnecessarily restrictive. To avoid this shortfall, the problem was addressed from the other end: without specifying the language, the author tried to understand the general semantics of the logic expressible through the evidential universes. Once the semantic was understood, the pool of suitable languages was narrowed, and the criteria for a propositional language definition became clearer.

Putting the priority on the semantics requires selecting a ground theory providing the direction of search. This ground theory must answer the fundamental questions about the nature of the mathematical objects. An important feature of the Dempster-Shafer theory is its ability to incorporate changes in the world. Viewing the universe as a dynamic system moves the understanding of the mathematical objects away from Russell-Wittgenstein paradigms. To achieve a reasonable balance between imprecise and rigorous, the analysis of the nature of mathematical objects was conducted from the viewpoint of constructive mathematics or Brouwer's intuitionism. Accepting a certain philosophical stance had formal implications. From the very beginning it was assumed that formula $p \vee \neg p$ is true only if one of the disjunction members is decidable.

Once the philosophical questions were addressed, the attention was switched to the formal part of the work. Given the considerations above, representing Dempster-Shafer frames of discernment through Kripke models was a natural choice. Even though semantic models were used for interpreting the Dempster-Shafer theory earlier, the proposed procedure is different. Only models that are both reflexive, transitive and antisymmetric were used. The set of formulae validated in the resulting models was analysed and shown to be equal to the 'minimal' intuitionist logic **Int** thus ensuring that the formalism is useful in a sense that underlying logic is complete and sound.

Another distinguishing feature of the proposed approach is its full parallelism with the Dempster-Shafer theory: while many earlier interpretations offered the methods for calculating the propositions' beliefs and have some kind of analogue to the evidence combination rule, they lack the parallels between frame transformations in the evidential setup and the logic constructions.

The all embracing nature of the developed procedure makes it possible to use the framework as reasoning apparatus in decision-making systems that base their inference either on imprecise or uncertain information. This possible application is well in line with the original inspiration behind the whole work which followed from the author's interest in decision-making agents and different formalisms used to rank

the alternatives. The next section provides a more detailed overview of the differences between the proposed approach and earlier interpretations.

5.2 Comparison with earlier interpretations

The procedure developed in this work offers a comprehensive approach to logic interpretation of the Dempster-Shafer theory. The earliest applications of the Dempster-Shafer theory used Boolean logic for inference. In this case only calculating the beliefs was different from the classical probability theory. The set of true propositions was determined only by the non-zero mass assignments. Such situation may be described by validating all the formulae true at a single node. Using a single-node semantic model does not capture the spirit of the Dempster-Shafer theory too well. Having a single node does not reflect the possibility to learn new facts and operate with different sets of beliefs at different moments of time. Using Boolean logic thus does not really reflect the state of affairs as it is implied by the frames of discernment.

The problem with Boolean inference was observed quite early and numerous interpretations that used richer semantics followed. Some of these results were reviewed in Chapter 1. The proposed framework differs from most of them. First of all, most of proposed interpretations, while accepting the need to have different semantic models in order to represent the Dempster-Shafer theory, focussed their attention on expressing the beliefs through modal operators. This premise led to constructing modal logics whose semantics were determined by the choice of connectives rather than by the underlying theory. As an example of such logics the already reviewed $L\Pi_{\frac{1}{2}}$ can be mentioned [76]. Operator-centred approach may lead to negative consequences as in [71] where the particular requirements to modal connective led to a logic that were not capable to provide an analogue of the Dempster-Shafer evidence combination rule.

The approach which is the most close in spirit to the approach in this work was used by Tsiporkova et al. [77], [78], who took the most general definition of the modal operator and looked into what properties followed from linking the Dempster-Shafer evidential setup and modal logics. The results presented by the authors include the procedure that induces beliefs and has some analogue of the Dempster-Shafer evidence combination rule. On the other hand, there is no analysis of the resulting set of semantic models, and the completeness properties of the logic are not explored.

Boutilier [75] takes the approach in which the set of modal connectives is defined based on the author's understanding of the evidential setup, and the main focus of the work is on analysing the semantic models. The modal models introduced by Boutilier

reflect the possibility to learn facts, and the accessibility relation is a total preorder. Therefore, the models introduced by Boutilier allow for *clusters* of possible worlds that are equally likely. Note that a total pre-order (A, \preceq) can be factorised into a total order $(A/\sim, <)$ (or a chain) by canonical projection $\pi : A \rightarrow A/\sim$, where A/\sim is the quotient set (set of equivalence classes $[a] \equiv \{b \in A : b \sim a\}$ and π is a map $a \mapsto [a]$).

In the proposed approach, relation R is a partial order (antisymmetric), which means that different nodes can only represent different worlds. A chain corresponding to A/\sim is a particular case of R (when it is a total order). In this sense, models induced by the proposed approach are richer.

The other significant difference between Boutilier's approach and the current approach is Boutilier's desire to develop qualitative rather than quantitative framework, which results in providing a procedure for getting a non-numerical ordering of the worlds rather than a developed apparatus for calculating the beliefs in propositions.

Most of the proposed frameworks do not pay much attention to studying the properties of induced semantic models. The author believes that studying the semantic models, especially in combinatorial or algebraic context, leads to interesting results and may provide useful tools for developing effective algorithms for the whole undertaking.

Early graphical representation of the Dempster-Shafer approach can be found in Barnett's work [79], [82] dating from the early eighties. The results are, however, not universal; the evidential setups are represented as binary trees that may be used to calculate the beliefs. The model only covers the setups described by separable support functions, and no semantic analysis is available as the authors developed their approach within the realm of Boolean logic as a reasoning apparatus.

None of the earlier approaches which are known to the author followed the principles of constructive mathematics, and none of the proposed logics explicitly ruled out *tertium non datur*. Instead, additional axioms were often introduced. Even though the situation when *tertium non datur* is explicitly refuted does not occur too often, the possibility to incorporate 'empty evidence' should be avoided as the example in Section 3.6 demonstrates.

Overall, the proposed approach may be described as a framework that addresses all the aspects of reasoning with beliefs: finding a set of true formulae, calculating the beliefs, and providing graphical representation. The approach can also be seen as the one that provides the *minimal set* of true formulae. As shown by Alechina the

set of true formulae validated in the induced semantic models is equal to **Int** [99].

5.3 Summary

The proposed approach represents a support function S over a Dempster-Shafer frame of discernment Θ as a semantic model \mathcal{M} with weighted nodes. The distinguishing feature of the approach is its ability to represent support functions and evidence combination through non-trivial semantic models.

The traditional approach to reasoning in the Dempster-Shafer universe is to use Boolean logic, which is semantically equivalent to all formulae valid at a single node. However, in many cases it leads to 'too many' formulae being valid.

An attractive property of non-trivial semantic models is their ability to *refute* certain formulae at different nodes. In the current work it was shown that the underlying logic, whose formulae are guaranteed to be valid at any node of any model is **Int**. Thus, any formula that is not in **Int** may be refuted at some node of a semantic model. It does not mean that given a formula and a model it is guaranteed that the formula will be refuted at some node. It only means that given a formula not in **Int** there is a model that refutes it. It is also possible to find formulae refuted by the model if the model is known.

The simplest example of the relation above is $p \vee (p \rightarrow \perp)$, and any model that has two connected nodes, instantiating p in a 'later' node leads to $p \vee (p \rightarrow \perp)$ being invalid at the earlier one. Depending on the model structure the set of formulae that is not valid at certain nodes changes. Knowing the valuation of particular formulae gives a possibility to limit the set of possible inferences at each particular node since there is a set of formulae $\phi \notin \mathbf{Int}$ that are not valid at that node.

In terms of building an inferential apparatus the limitations above mean that only logically possible conclusions are made based on each belief (the nodes of a model represent beliefs).

Let us now return to the example given in Section 3.3.3. Recall that the support function described a sensor checking the colour of some faraway object. Now let us try to make inference based on measurements of many such sensors. We also need to take into account a belief that some of them may be broken without any opportunity to check if they are indeed broken: the sensors are not accessible for examination (e.g. they are on Mars).

There are a few points that should be considered in a situation like that.

1. The difference between probabilistic/statistical approach and the Dempster-

Shafer approach — one cannot really formulate and test a statistical hypothesis. It is not an experiment that may be reproduced, thus a probability law describing the odds that the sensor is broken is not available, because we cannot formulate the requirements for a test to achieve desirable statistical significance. In this case, beliefs are, perhaps, a more appropriate concept, even though it is a matter of opinion. The Dempster-Shafer theory is a computational formalism to deal with beliefs, not probabilities.

2. The set of possible values of the parameter x measured by the sensor is Θ . Select a relevant subset of Θ , where mass assignments are not zero and build a semantic model with mass assignments according to Definitions 3.3.2, 3.3.3 and 3.3.4. Possible inference at each node of this model will be different (i.e. sets of provable formulae are different). Using models that represent logic *Int* allows one to operate within a richer semantic. In case of Boolean logic the model is trivial. With *Int* there is a possibility of different nodes that represent different possible worlds that could correspond to the same data, and which could produce different inference. The mass assignments provide some numerical estimate how believable each world is.
3. To illustrate the previous point consider the following situation. Assume that temperature of red objects may be measured, but that the colour sensor cannot measure temperature, so that some other, difficult to operate, sensor must be used. The temperature of blue and white objects is irrelevant. In case of Boolean semantics there is nothing that prevents a decision maker from evaluating formulae that involve $p \vee (p \rightarrow \perp)$, where $p = \text{'object is hot'}$ even at the nodes that correspond to, say, white objects. Moreover, according to the axioms of Boolean logic, the agent must use this formula as a tautology even in cases when there is no information about the temperature of the object and, strictly speaking, using *any* proposition involving it, is not justified.
4. There could be too many possible worlds, and the problem of proving/validating too many formulae. In the proposed approach, the nodes represent only non-trivial beliefs (i.e. the ones that are neither one or zero), and this refutes certain formulae from the very beginning. Having numerical mass assignments on the nodes can further help to prioritize the inference process by focussing on the most 'believable' worlds rather than proving all possible formulae. This ability may be quite important for computational applications.

5. The decision can be made whether the belief in some state of affairs (world) is strong enough or not and some additional measurement have to be made.
6. New evidence can be obtained, and it can change the beliefs. The process of incorporating new evidence results in updating the models as well, and the procedures for it were developed (refinement, coarsening) in Chapter 4. Thus, after obtaining and incorporating new evidence, the new inference process may reuse many results of previous computations.

The possibility to limit the number of valid formulae at a particular node also suggests a direction for future research that was analysed in some detail in Section 3.5. Even though there are no explicit conditions that the semantic models representing support functions should satisfy, some properties can be inferred from the evidential setup. It was shown that having at least two connected nodes leads to refuting $p \vee (p \rightarrow \perp)$, there are similar conditions on the presence of chains and so on. Many intermediate logics are defined through conditions on frames which can be applicable to particular models representing support functions. Determining which conditions on frames are satisfied by a particular semantic model will help the decision maker to limit the set of valid formulae and thus lead to more effective inference.

In the most general sense the proposed approach agrees with the ideas behind the rule based systems that operate using probabilistic logic networks [100]. The latter is a much more ambitious work aimed at providing both fundamental theory and implementation of the authors' ideas about cognitive development which they credit to works of Piaget and Vygotsky. The same authors discuss the aspects of the uncertain inference and come to the conclusion that one of the hardest tasks is to control preference with regard to which rules apply first and what should wait until more knowledge is available:

The subtlest part of uncertain inference is inference control: the choice which inferences to do, in what order [101].

The semantic models contribute to addressing the problem outlined above, by giving a natural hierarchy of the rules based on the information at disposal of a decision maker at a particular state of the world. When viewed from this angle of view, the proposed formalism is especially attractive. As it is observed in the paper just quoted, people are not very efficient in applying inference rules, but they exceed machines in their ability to *choose* these rules. The Dempster-Shafer theory is a formalism aimed at representing human beliefs rather than additive probabilities. It can model human

behaviour when the probability theory is not applicable. Thus a Dempster-Shafer theory based mechanism for inference rule selection can be an alternative approach.

Bibliography

- [1] Glenn Shafer. *Allocations of Probability: A Theory of Partial Belief*. PhD thesis, Princeton University, 1973.
- [2] Nick Bezhanishvili. *Lattices of Intermediate and Cylindric Modal Logics*. PhD thesis, Institute for Logic, Language, and Computation, Universiteit van Amsterdam, 2006.
- [3] Zuzana Haniková. *Mathematical and Metamathematical Properties of Fuzzy Logic*. PhD thesis, Charles University, Prague, 2007.
- [4] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, 81:23–69, 2003.
- [5] A. N. Kolmogorov. *Foundations of the Theory of Probability*, Second English Edition. Chelsea Publishing Company, New York, 1956.
- [6] George J. Klir. Uncertainty-based information: a critical review. In Vilém Novák and Irina Perfilieva, editors, *Discovering the world with Fuzzy logic*, pages 29–57. Physica Verlag, 2000.
- [7] George J. Klir and Mark J. Wierman. *Uncertainty- Based Information: elements of generalized information theory*. Physica-Verlag, 1998.
- [8] D. Ellsberg. Risk, ambiguity and the Savage axioms. *Quart. J. Econom.*, 75:643–669, 1961.
- [9] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [10] D. Schmeidler. Integral representation without additivity. *Proc. of AMS*, 97(2):255–261, 1986.

- [11] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.
- [12] P. Wakker. A behavioral foundation for fuzzy measures. *Fuzzy sets and systems*, 37:327–350, 1990.
- [13] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [14] A. P. Dempster. A generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society*, 30:205–247, 1968.
- [15] G.-C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Z. fr Wahrscheinlichkeitsth*, 2:340–368, 1964.
- [16] M. Grabisch. The interaction and Möbius representations of fuzzy measures on finite spaces, k -additive measures: a survey. In M. Grabisch, T. Murofushi, M. Sugeno, and J. Kacprzyk, editors, *Fuzzy Measures and Integrals - Theory and Applications*, pages 70–93. Physica Verlag, 2000.
- [17] Kari Sentz and Scott Ferson. Combination of evidence in Dempster-Shafer theory. Technical report, Sandia National Laboratories, 2002.
- [18] Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic-Studia Logica Library*. Kluwer Academic Publishers, 1998.
- [19] Aristotle. *Categories*, translated by E. M. Edghill. eBooks@Adelaide, University of Adelaide, Australia, 2007.
- [20] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
- [21] N.N. Morsi. Categories of fuzzy topological spaces with localization at points. In *28th Linz seminar on Fuzzy set theory*, 2008.
- [22] J. Harding, C. Walker, and E. Walker. On some topological properties of the subalgebra of normal convex functions of the algebra of truth values of type-2 fuzzy sets. In *28th Linz seminar on Fuzzy set theory*, 2008.
- [23] W. Pedrycz and F. Gomido. *An Introduction to fuzzy sets: Analysis and Design*. MIT Press, Cambridge, Massachusetts, 1998.

- [24] Philippe Smets. Imprecision and uncertainty. In J.-E. Dubois and N. Gershon, editors, *Modeling Complex Data for Creating Information*, pages 39–47. Springer, 1996.
- [25] George J. Klir and Tina A. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice-Hall International, Englewood Cliffs, New Jersey, 1988.
- [26] W. Pedrycz. Fuzzy sets in pattern recognition: Methodology and methods. *Pattern Recognition*, 23(1/2):121–146, 1990.
- [27] Soodamani Ramalingam. *Fuzzy Paradigm for 3D Object Recognition*. PhD thesis, University of Melbourne, Australia, 1997.
- [28] R. Soodamani and Z.Q. Liu. A fuzzy hough transform approach to shape description. *Intl. Journal of Image and Graphics*, 2(4):603–616, 2002.
- [29] S. Ramalingam, Z.Q. Liu, and D. Iourinski. Curvature-based fuzzy surface classification. *IEEE Trans Fuzzy Systems*, 14:573–589, 2006.
- [30] K. Miyajima and A. Ralescue. Modeling of natural objects including fuzziness and application to image understanding. In *Proceedings of the IEEE*, pages 1049–1054, 1993.
- [31] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5:131–296, 1953.
- [32] L.A. Zadeh. Probability measures of fuzzy events. *J. Math. Analysis and Appl.*, 23:421–427, 1968.
- [33] H.L. Royden. *Real Analysis*. Macmillan, 1988.
- [34] Toshiaki Murofushi and Michio Sugeno. An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy Sets and Systems*, Volume 29 , Issue 2:201 – 227, 1989.
- [35] Ulrich Höhle. Integration with respect to fuzzy measures. In *Proc. IFAC Symp. on Theory and Appl of Digital Control, New Delhi*, pages 35–37, 1982.
- [36] M. Sugeno. *Theory of Fuzzy Integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
- [37] S. Weber. \perp -decomposable measures and integrals for archimedean t-conorms \perp . *J. Math. Anal. Appl.*, 101:114–138, 1984.

- [38] T. Murofushi and M. Sugeno. Fuzzy t-conorm integrals with respect to fuzzy measures: generalization of Sugeno integral and Choquet integral. *Fuzzy Sets & Systems*, 42, 1991.
- [39] M. Grabish, H. Nguyen, and E. Walker. *Fundamentals of uncertainty calculi with applications to fuzzy inference*. Kluwer academic publishers, Dordrecht, 1995.
- [40] Petr Hájek. On logics of approximate reasoning. In Michael Mazuch and László Pólos, editors, *Knowledge Representation and Reasoning under Uncertainty*, volume 808 of *Lecture Notes in Computer Science*, pages 15–27. Springer, 1994.
- [41] J. Lukasiewicz. *Selected works by Jan Lukasiewicz*, chapter O logice trójwartościowej (in Polish). *Ruch filozoficzny* 5:170171. English translation: On three-valued logic, page 8788. North Holland, Amsterdam, 1970.
- [42] Jan Lukasiewicz. *Elements of Mathematical Logic*. New York, Macmillan., 1964. Translated from Polish by Olgierd Wojtasiewicz.
- [43] Erich Peter Klement, Radko Mesiar, and Endre Pap. *Triangular Norms*. Serie Trends in Logic, Studia Logica. Kluwer Academic Publishers, Dordrecht/Boston/London, 2000.
- [44] Erich Peter Klement. Triangular norms - what they are and what they are good for. In *2nd IEEE-SOFA workshop on soft computing applications, Oradea-Romania, Gyula-Hungary*, 2007.
- [45] Janos Fodor and Imre J. Rudas. On a class of continuous triangular norms. In *2nd IEEE-SOFA workshop on soft computing applications, Oradea-Romania, Gyula-Hungary*, 2007.
- [46] János Fodor. Left-continuous t-norms in fuzzy logic: An overview. *Acta Polytechnica Hungarica*, 1(2), 2004.
- [47] S. Saminger-Platz, E.P. Klement, and R. Mesiar. Bounded lattices, complete sublattices, and extensions of t-norms. In *29th Linz seminar on Fuzzy set theory*, 2008.
- [48] Alexander Chagrov and Michael Zakharyashev. *Modal Logic*, volume 35 of *Oxford Logic Guides*. Oxford Science Publications, 1997.

- [49] Jean-Yves Girard. Truth, modality and intersubjectivity. *MSCS: Mathematical Structures in Computer Science*, 17:1153–1168, 2007.
- [50] Dov M. Gabbay, Ian Hodkinson, and Mark Reynolds. *Temporal Logic*, volume 28 of *Oxford Logic Guides*. 1994.
- [51] Vincent Hendricks and John Symons. Epistemic logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2006.
- [52] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge tracts in theoretical computer science*. Cambridge University Press, 2001.
- [53] Robert Goldblatt. *Mathematics of Modality*, volume 43 of *CSLI lecture notes*. Hemel Hempstead, 1993.
- [54] L. E. J. Brouwer. *Collected Works 1, Philosophy and Foundations of Mathematics*. Amsterdam, 1975.
- [55] L. E. J. Brouwer. *Collected works*. North-Holland Publishing Co., 1975-6.
- [56] L. E. J. Brouwer. *Brouwer's Cambridge lectures on intuitionism; edited by D. Van Dalen*. Cambridge University Press, 1981.
- [57] L.E.J. Brouwer. Historical background, Principles and Methods of Intuitionism. *South African Journal of Science*, 49:139–146, 1952.
- [58] A. Heyting. *Intuitionism: an Introduction*. Amsterdam, 3rd edition, 1972.
- [59] D. S. Bridges and F. Richman. *Varieties of Constructive Mathematics*. Cambridge, 1987.
- [60] P.S. Novikov. *Constructive Mathematical Logic from the Point of View of Classical Logic*. Nauka, Moscow, 1977.
- [61] Grigori Mints. *A short introduction to intuitionistic Logic*. Kluwer Academic/Plenum Publishers, 2000.
- [62] Dirk van Dalen. Intuitionistic logic. In D. Gabbay and F. Guenther, editors, *Handbook of philosophical logic*, volume III, pages 225–340. D. Reidel publishing company, 1986.

- [63] Paul R. Halmos. *Algebraic Logic*. Chelsea Publishing Company, New York, USA, 1962.
- [64] Mac Lane Saunders. *Categories for the Working Mathematician*. Springer, 1998.
- [65] M. Dummett. *Elements of Intuitionism*, volume 39 of *Oxford Logic Guides*. Oxford Science Publications, 2nd edition, 2000.
- [66] Sergei S. Goncharov. *Countable Boolean Algebras and Decidability*. Siberian School of Algebra and Logic. Consultants Bureau, NY, London, Moscow, 1997.
- [67] S.C. Kleene. *Introduction to Metamathematics*, volume 1 of *Bibliotheca Mathematica*. North-Holland, 1952.
- [68] A. Tarski. *Logic, Semantics, Metamathematics*. Clarendon Press, Oxford, 1956.
- [69] Saul Kripke. *Naming and Necessity*. Harvard University Press, 1980.
- [70] Arif Ahmed. *Saul Kripke*. Continuum, 2007.
- [71] Petr Hájek and Dagmar Harmancová. A note on Dempster rule. January 2003.
- [72] Philippe Smets. What is Dempster-Shafer model? In R. R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in Dempster-Shafer theory of evidence*, pages 5–34. Wiley, 1994.
- [73] Didier Dubois, Jrme LANG, and Henri PRADE. Timed possibilistic logic. In *Handbook of logic in Artificial Intelligence and logic programming*, pages 439–513. Oxford University Press, 1991.
- [74] Luis Fariñas del Cerro and Andreas Herzig. A modal analysis of possibility theory. In *FAIR*, pages 11–18, 1991.
- [75] C. Boutilier. Modal logic for qualitative possibility and beliefs. In Didier Dubois, Michael P. Wellman, Bruce D'Ambrosio, and Phillipe Smets, editors, *Proceedings of the eighth conference on Uncertainty in Artificial Intelligence*, pages 17–24. Morgan-Kaufmann Publihers, 1992.
- [76] L. Godo, P. Hájek, and F. Esteva. A fuzzy modal logic for belief functions. *Fundamenta Informaticae*, 57:127–146, 2003.

- [77] Veselka Boeva, Elena Tsiporkova, and Bernard De Baets. Plausibility and belief measures induced by Kripke models. In *ISIPTA*, pages 44–50, 1999.
- [78] Elena Tsiporkova, Bernard De Baets, and Veselka Boeva. Dempster-Shafer theory framed in modal logic. *International journal of approximate reasoning*, 21(2):157–175, 1999.
- [79] J.A. Barnett. Computational methods for a mathematical theory of evidence. In *7th international joint conference on AI(IJCAI)*, pages 868–875, 1981.
- [80] J. Gordon and E.H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.
- [81] Glenn Shafer and R. Logan. Implementing Dempster rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987.
- [82] J.W. Guan and D.A. Bell. *Evidence theory and its applications*, volume 2. North-Holland, 1992.
- [83] Dmitri Iourinski. *Views on Fuzzy Sets and Systems from Different Perspective. Philosophy, Logic; Criticisms and Applications*, chapter Phenomenology as criterion for formalism choice. Studies in Fuzziness and Soft Computing. Berlin [et al.]: Springer, 2009.
- [84] Dmitri Iourinski and Je X. For. Semantical models for evidential frameworks. In *2nd IEEE-SOFA workshop on soft computing applications, Oradea-Romania, Gyula-Hungary*, 2007.
- [85] John P. Burgess. Kripke models. In Alan Berger, editor, *Saul Kripke*, Contemporary Philosophers. Cambridge University Press, preprint.
- [86] Paul Halmos and Steven Givant. *Logic as Algebra*, volume 21 of *Dolciani Mathematical Expositions*. The Mathematical Association of America, 1998.
- [87] Richard A. Brualdi and H.J. Ryser. *Combinatorial Matrix Theory*. Cambridge University Press, 1991.
- [88] G. Boole. *Mathematical Analysis of Logic. Being an Essay towards a Calculus of Deductive Reasoning*. Barclay & Macmillian, London, 1847.

- [89] G. Frege. Begriffsschrift. In J. van Heijenoort, editor, *From Frege to Gödel: A source Book in Mathematical Logic, 1879-1931*, pages 5–82. Harvard University Press.
- [90] B. Russell and A.N. Whitehead. *Principia Mathematica*. Cambridge University Press, 1910.
- [91] J. M. Dunn and G. M. Hardegree. *Algebraic Methods in Philosophical Logic*, volume 41 of *Oxford Logic Guides*. Oxford Science Publications, 2001.
- [92] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Routledge & Kegan Paul, London, 1961.
- [93] M. van Atten. Luitzen Egbertus Jan Brouwer. In *The Stanford Encyclopedia of Philosophy*. 2003.
- [94] Edmund Husserl. *Analyses concerning passive and active synthesis: lectures on transcendental logic*. Dordrecht, London: Kluwer Academic Publishers, 2001.
- [95] Mark van Atten. *Brouwer meets Husserl: on the phenomenology of choice sequences*, volume 335 of *Synthese library*. Springer, 2007.
- [96] J.W. Guan and D.A. Bell. *Evidence theory and its applications*, volume 1. North-Holland, 1991.
- [97] P. Blackburn and J. van Benthem. Modal logic: A semantic perspective. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *The Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*. Elsevier, 2006.
- [98] Wikipedia. Dempster-Shafer Theory – Wikipedia, the free encyclopedia, accessed 2008.
- [99] Natasha Alechina. Note about embedding L_{DS} into Int. Personal communication, 19 February 2009.
- [100] Ben Goertzel, Matthew Iklé, Izabela Lyon Freire Goertzel, and Ari Heljakka. *Probabilistic Logic Networks: A Comprehensive Conceptual, Mathematical and Computational Framework for Uncertain Inference*. Springer, 2008.
- [101] Ben Goertzel and Stephan Vladimir Bugaj. Stages of cognitive development in uncertain-logic based AI systems. In *AGIRI workshop*. Washington, DC, 2006.